Collins, C.
10/768511
Seq. IDs 5 & 8


```
OM nucleic - nucleic search, using sw model

Run on:         December  6, 2005, 01:29:10 ; Search time 14083 Seconds
                                          (without alignments)
                                          17549.906 Million cell updates/sec

Title:          US-10-768-511-5
Perfect score:  4348
Sequence:       1 atggcgcgccgcactcacgt..........gcgcaattcccagagctcgc 4348

Scoring table:  IDENTITY_NUC
                Gapop 10.0 , Gapext 1.0

Searched:       5883141 seqs, 28421725653 residues

Total number of hits satisfying chosen parameters:      11766282

Minimum DB seq length: 0
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%
                 Maximum Match 100%
                 Listing first 45 summaries

Database :       GenEmbl:*
                1:  gb_ba:*
                2:  gb_in:*
                3:  gb_env:*
                4:  gb_om:*
                5:  gb_ov:*
                6:  gb_pat:*
                7:  gb_ph:*
                8:  gb_pr:*
                9:  gb_ro:*
               10:  gb_sts:*
               11:  gb_sy:*
               12:  gb_un:*
               13:  gb_vi:*
               14:  gb_htg:*
               15:  gb_pl:*
```

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

	Result No.	Score	% Query Match	Length	DB	ID	Description
	1	4348	100.0	4348	6	AR490068	AR490068 Sequence
	2	4348	100.0	4348	6	AX281543	AX281543 Sequence
	3	673	15.5	723	6	AR490065	AR490065 Sequence
	4	673	15.5	723	6	AX281540	AX281540 Sequence
	5	520.8	12.0	3667	15	SPCDC21	X58824 Schizosacch
c	6	492.4	11.3	153751	2	AC116551	AC116551 Dictyoste
	7	480.2	11.0	2733	6	AX489180	AX489180 Sequence
c	8	471.4	10.8	110000	14	CR954214_1	Continuation (2 of
	9	470.8	10.8	5392	15	SCU14731	U14731 Saccharomyc
	10	470.8	10.8	27166	15	SC9531X	Z49919 S.cerevisia
	11	470.8	10.8	165536	15	SCCHRXVI	Z71255 S.cerevisia
c	12	464.4	10.7	110000	15	CR380957_02	Continuation (3 of
	13	458.6	10.5	110000	15	CR382132_07	Continuation (8 of
	14	457.8	10.5	110000	15	AE016817_06	Continuation (7 of
c	15	456	10.5	110000	15	CR382123_15	Continuation (16 o
	16	444.2	10.2	3290	5	BC074670	BC074670 Xenopus t
	17	436.8	10.0	3048	5	BC072870	BC072870 Xenopus l
	18	435.2	10.0	2970	5	XLU46131	U46131 Xenopus lae
	19	435.2	10.0	3051	5	XLU29178	U29178 Xenopus lae
	20	430.8	9.9	3273	6	CQ981174	CQ981174 Sequence
	21	430.8	9.9	3273	6	AX333086	AX333086 Sequence
	22	430.8	9.9	3273	6	AX333087	AX333087 Sequence
	23	430.8	9.9	3273	6	AX333494	AX333494 Sequence
	24	430.8	9.9	3273	8	HSP1CDC21	X74794 H.sapiens P
	25	430.8	9.9	3273	10	G28594	G28594 human STS S
	26	430.4	9.9	3394	6	AX281814	AX281814 Sequence
	27	430	9.9	2893	5	BC083031	BC083031 Xenopus l
	28	429.2	9.9	2851	8	BC031061	BC031061 Homo sapi
	29	427.6	9.8	2860	6	BD159700	BD159700 Primer fo
	30	427.6	9.8	2860	6	AX882395	AX882395 Sequence
	31	427.6	9.8	2860	8	AK022899	AK022899 Homo sapi
	32	425.2	9.8	3087	5	XLU44049	U44049 Xenopus lae
	33	413	9.5	110000	15	CR382137_13	Continuation (14 o
	34	407.8	9.4	3042	5	BC065958	BC065958 Danio rer
	35	407.6	9.4	2556	6	CQ717883	CQ717883 Sequence
	36	395.8	9.1	3040	5	BC056514	BC056514 Danio rer
	37	392.8	9.0	3221	9	MUSMCDC21	D26089 Mus musculu
	38	392.8	9.0	3283	9	AK220442	AK220442 Mus muscu
	39	391.2	9.0	3360	9	BC013094	BC013094 Mus muscu
	40	384.6	8.8	244538	14	AC129365	AC129365 Rattus no
	41	381.6	8.8	2776	2	AK114961	AK114961 Ciona int
	42	371.6	8.5	241140	14	AC119094	AC119094 Rattus no
	43	329.6	7.6	2776	2	S80255	S80255 dpa=disc pr
	44	329	7.6	2822	2	BT021329	BT021329 Drosophil
	45	328.8	7.6	40751	15	SPCC24B10	AL157991 S.pombe c

ALIGNMENTS

```
RESULT 1
AR490068
LOCUS       AR490068                4348 bp    DNA     linear   PAT 15-MAY-2004
DEFINITION  Sequence 5 from patent US 6710229.
ACCESSION   AR490068
VERSION     AR490068.1  GI:47257191
KEYWORDS    .
SOURCE      Unknown.
  ORGANISM  Unknown.
            Unclassified.
REFERENCE   1  (bases 1 to 4348)
  AUTHORS   da Costa e Silva,O., Bohnert,H.J., van Thielen,N., Chen,R. and
            Sarria-Millan,R.
  TITLE     Cell cycle stress-related proteins and methods of use in plants
  JOURNAL   Patent: US 6710229-A 5 23-MAR-2004;
            BASF Plant Science GmbH; Ludwigshafen;
            WOX;
FEATURES             Location/Qualifiers
     source          1..4348
                     /organism="unknown"
                     /mol_type="genomic DNA"
ORIGIN

  Query Match            100.0%;  Score 4348;  DB 6;  Length 4348;
  Best Local Similarity  100.0%;  Pred. No. 0;
  Matches 4348;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy          1 ATGGCGCGCCGCACTCACGTGAGGAATTGCACCTCCTTGTTCTGCGACGGTTCCATTCTT 60
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db          1 ATGGCGCGCCGCACTCACGTGAGGAATTGCACCTCCTTGTTCTGCGACGGTTCCATTCTT 60

Qy         61 TTTGGTTTTTAGTTTGCAAATCTTGATCGTGGAGTTGAGAAAAAGGGCGGTTCGTTGTCT 120
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db         61 TTTGGTTTTTAGTTTGCAAATCTTGATCGTGGAGTTGAGAAAAAGGGCGGTTCGTTGTCT 120

Qy        121 TGAGGTGTTCTTGTTGATTGTTCGTCATGGAAAATAATGATGCACTTGACATTGGAGCCG 180
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        121 TGAGGTGTTCTTGTTGATTGTTCGTCATGGAAAATAATGATGCACTTGACATTGGAGCCG 180

Qy        181 TGTCGTCCCCATATCCTTCGCAATCTGAAGGAGTGTCTACGCCATTGCCGCAAGTAACAT 240
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
```

```
Db        181 TGTCGTCCCCATATCCTTCGCAATCTGAAGGAGTGTCTACGCCATTGCCGCAAGTAACAT 240

Qy        241 CACCGAGCTTCGACAATGCAGCCTCACCCGTGGCCGGGCGGAGGGCCGTACGGCAGACCC 300
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        241 CACCGAGCTTCGACAATGCAGCCTCACCCGTGGCCGGGCGGAGGGCCGTACGGCAGACCC 300

Qy        301 CTACATCTGCAGTTCGAAGGAGAGGGAGAGAAACGGATTCCGCTCGTCGTAGGAGGAGTC 360
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        301 CTACATCTGCAGTTCGAAGGAGAGGGAGAGAAACGGATTCCGCTCGTCGTAGGAGGAGTC 360

Qy        361 GATCTCGCAGTTTAGGCAATTCTGTTTATAGTTCCCCTTACGATGCGGGGACTCCTGGAA 420
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        361 GATCTCGCAGTTTAGGCAATTCTGTTTATAGTTCCCCTTACGATGCGGGGACTCCTGGAA 420

Qy        421 CTCCTGGAACTCCAGTGGCTACTCCGGTTTACGCTACCCCAGTCGGTACACCTATGGGTA 480
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        421 CTCCTGGAACTCCAGTGGCTACTCCGGTTTACGCTACCCCAGTCGGTACACCTATGGGTA 480

Qy        481 CCCCATCGTTCCATCGTGGCACGCCACAGTACAAACAGCGCAGTGAGCTTGGTTCCCAGG 540
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        481 CCCCATCGTTCCATCGTGGCACGCCACAGTACAAACAGCGCAGTGAGCTTGGTTCCCAGG 540

Qy        541 GGAAGCCTCTACATCGGAGACGTCGATCTCAATCCAGAGAACCCGGGCATCGATCTCCTT 600
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        541 GGAAGCCTCTACATCGGAGACGTCGATCTCAATCCAGAGAACCCGGGCATCGATCTCCTT 600

Qy        601 CAAGGGAACCTAGTGCTGATGGGCGTCCCTCTGAATCTGCTGAGCCAGATGACACTTTGG 660
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        601 CAAGGGAACCTAGTGCTGATGGGCGTCCCTCTGAATCTGCTGAGCCAGATGACACTTTGG 660

Qy        661 GTGGAGAATATGCTTATGTTTGGGGGACGAATGTTAACATTCCAGATGTGCTTAGGGCGA 720
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        661 GTGGAGAATATGCTTATGTTTGGGGGACGAATGTTAACATTCCAGATGTGCTTAGGGCGA 720

Qy        721 TTCGTCGATTTCTCCACAATTATCGTTCGAGTGCTCATGATCTTAATTCCAAGTACATCC 780
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        721 TTCGTCGATTTCTCCACAATTATCGTTCGAGTGCTCATGATCTTAATTCCAAGTACATCC 780

Qy        781 AGATCATAGAGGAGACTGTGGAGCGTGAGGAGGATACTCTAAATATCGACATGTCAGACA 840
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        781 AGATCATAGAGGAGACTGTGGAGCGTGAGGAGGATACTCTAAATATCGACATGTCAGACA 840

Qy        841 TTTATGACCATGATCCTGATCTATACGCAAAAATTGTTCGATACCCACTCGACATCATCC 900
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        841 TTTATGACCATGATCCTGATCTATACGCAAAAATTGTTCGATACCCACTCGACATCATCC 900

Qy        901 CCCTGTTGGACACTGAGTGTCAGGAAGTTGCTACCTCTTTACTACCAACGTTTGAGAAGC 960
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        901 CCCTGTTGGACACTGAGTGTCAGGAAGTTGCTACCTCTTTACTACCAACGTTTGAGAAGC 960

Qy        961 ATATTGAGGCCAGACCTTTCAATCTCAAAGCATCGGTGCACATGCGTGAACTCAACCCTT 1020
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        961 ATATTGAGGCCAGACCTTTCAATCTCAAAGCATCGGTGCACATGCGTGAACTCAACCCTT 1020

Qy       1021 CAGATATAGACAAATTGGTTTCTGTTAAAGGAATGGTTATCCGGTGCAGTTCTATCATAC 1080
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1021 CAGATATAGACAAATTGGTTTCTGTTAAAGGAATGGTTATCCGGTGCAGTTCTATCATAC 1080

Qy       1081 CTGAAATTAAGGGGGCCTTCTTCAAATGTTTAGTGTGTGGTCACTCGCCTCCGCTAGTTA 1140
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1081 CTGAAATTAAGGGGGCCTTCTTCAAATGTTTAGTGTGTGGTCACTCGCCTCCGCTAGTTA 1140

Qy       1141 CAGTTGTTAAAGGGCGGGTTGAGGAGCCAACAAGGTGTGAAAAGCCAGAATGTGCAGCAC 1200
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1141 CAGTTGTTAAAGGGCGGGTTGAGGAGCCAACAAGGTGTGAAAAGCCAGAATGTGCAGCAC 1200

Qy       1201 GGAATGCTATGTCTCTTATTCACAATCGATGCACTTTTGCAAATAAGCAGATAGTGCGTC 1260
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1201 GGAATGCTATGTCTCTTATTCACAATCGATGCACTTTTGCAAATAAGCAGATAGTGCGTC 1260

Qy       1261 TTCAAGAAACTCCAGATGCCATTCCTGAAGGAGAGACTCCACACACAGTCAGCATGTGTT 1320
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1261 TTCAAGAAACTCCAGATGCCATTCCTGAAGGAGAGACTCCACACACAGTCAGCATGTGTT 1320

Qy       1321 TATACAACACTATGGTTGATGCTGTGAAGCCTGGAGATCGTATTGAGGTAACAGGAGTTT 1380
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1321 TATACAACACTATGGTTGATGCTGTGAAGCCTGGAGATCGTATTGAGGTAACAGGAGTTT 1380

Qy       1381 TCAAGGCCATGGCAGTTCGAGTTGGTCATGGCGCGCCGCACTCACGTGAGGAATTGCACC 1440
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1381 TCAAGGCCATGGCAGTTCGAGTTGGTCATGGCGCGCCGCACTCACGTGAGGAATTGCACC 1440

Qy       1441 TCCTTGTTCTGCGACGGTTCCATTCTTTTTGGTTTTTAGTTTGCAAATCTTGATCGTGGA 1500
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1441 TCCTTGTTCTGCGACGGTTCCATTCTTTTTGGTTTTTAGTTTGCAAATCTTGATCGTGGA 1500

Qy       1501 GTTGAGAAAAAGGGCGGTTCGTTGTCTTGAGGTGTTCTTGTTGATTGTTCGTCATGGAAA 1560
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1501 GTTGAGAAAAAGGGCGGTTCGTTGTCTTGAGGTGTTCTTGTTGATTGTTCGTCATGGAAA 1560

Qy       1561 ATAATGATGCACTTGACATTGGAGCCGTGTCGTCCCCATATCCTTCGCAATCTGAAGGAG 1620
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1561 ATAATGATGCACTTGACATTGGAGCCGTGTCGTCCCCATATCCTTCGCAATCTGAAGGAG 1620

Qy       1621 TGTCTACGCCATTGCCGCAAGTAACATCACCGAGCTTCGACAATGCAGCCTCACCCGTGG 1680
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1621 TGTCTACGCCATTGCCGCAAGTAACATCACCGAGCTTCGACAATGCAGCCTCACCCGTGG 1680

Qy       1681 CCGGGCGGAGGGCCGTACGGCAGACCCCTACATCTGCAGTTCGAAGGAGAGGGAGAGAAA 1740
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1681 CCGGGCGGAGGGCCGTACGGCAGACCCCTACATCTGCAGTTCGAAGGAGAGGGAGAGAAA 1740

Qy       1741 CGGATTCCGCTCGTCGTAGGAGGAGTCGATCTCGCAGTTTAGGCAATTCTGTTTATAGTT 1800
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1741 CGGATTCCGCTCGTCGTAGGAGGAGTCGATCTCGCAGTTTAGGCAATTCTGTTTATAGTT 1800

Qy       1801 CCCCTTACGATGCGGGGACTCCTGGAACTCCTGGAACTCCAGTGGCTACTCCGGTTTACG 1860
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1801 CCCCTTACGATGCGGGGACTCCTGGAACTCCTGGAACTCCAGTGGCTACTCCGGTTTACG 1860

Qy       1861 CTACCCCAGTCGGTACACCTATGGGTACCCCATCGTTCCATCGTGGCACGCCACAGTACA 1920
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1861 CTACCCCAGTCGGTACACCTATGGGTACCCCATCGTTCCATCGTGGCACGCCACAGTACA 1920

Qy       1921 AACAGCGCAGTGAGCTTGGTTCCCAGGGGAAGCCTCTACATCGGAGACGTCGATCTCAAT 1980
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1921 AACAGCGCAGTGAGCTTGGTTCCCAGGGGAAGCCTCTACATCGGAGACGTCGATCTCAAT 1980

Qy       1981 CCAGAGAACCCGGGCATCGATCTCCTTCAAGGGAACCTAGTGCTGATGGGCGTCCCTCTG 2040
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1981 CCAGAGAACCCGGGCATCGATCTCCTTCAAGGGAACCTAGTGCTGATGGGCGTCCCTCTG 2040

Qy       2041 AATCTGCTGAGCCAGATGACACTTTGGGTGGAGAATATGCTTATGTTTGGGGGACGAATG 2100
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       2041 AATCTGCTGAGCCAGATGACACTTTGGGTGGAGAATATGCTTATGTTTGGGGGACGAATG 2100

Qy       2101 TTAACATTCCAGATGTGCTTAGGGCGATTCGTCGATTTCTCCACAATTATCGTTCGAGTG 2160
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       2101 TTAACATTCCAGATGTGCTTAGGGCGATTCGTCGATTTCTCCACAATTATCGTTCGAGTG 2160

Qy       2161 CTCATGATCTTAATTCCAAGTACATCCAGATCATAGAGGAGACTGTGGAGCGTGAGGAGG 2220
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       2161 CTCATGATCTTAATTCCAAGTACATCCAGATCATAGAGGAGACTGTGGAGCGTGAGGAGG 2220

Qy       2221 ATACTCTAAATATCGACATGTCAGACATTTATGACCATGATCCTGATCTATACGCAAAAA 2280
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       2221 ATACTCTAAATATCGACATGTCAGACATTTATGACCATGATCCTGATCTATACGCAAAAA 2280

Qy       2281 TTGTTCGATACCCACTCGACATCATCCCCCTGTTGGACACTGAGTGTCAGGAAGTTGCTA 2340
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       2281 TTGTTCGATACCCACTCGACATCATCCCCCTGTTGGACACTGAGTGTCAGGAAGTTGCTA 2340

Qy       2341 CCTCTTTACTACCAACGTTTGAGAAGCATATTGAGGCCAGACCTTTCAATCTCAAAGCAT 2400
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       2341 CCTCTTTACTACCAACGTTTGAGAAGCATATTGAGGCCAGACCTTTCAATCTCAAAGCAT 2400
```

```
Qy        2401 CGGTGCACATGCGTGAACTCAACCCTTCAGATATAGACAAATTGGTTTCTGTTAAAGGAA 2460
               ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        2401 CGGTGCACATGCGTGAACTCAACCCTTCAGATATAGACAAATTGGTTTCTGTTAAAGGAA 2460

Qy        2461 TGGTTATCCGGTGCAGTTCTATCATACCTGAAATTAAGGGGGCCTTCTTCAAATGTTTAG 2520
               ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        2461 TGGTTATCCGGTGCAGTTCTATCATACCTGAAATTAAGGGGGCCTTCTTCAAATGTTTAG 2520

Qy        2521 TGTGTGGTCACTCGCCTCCGCTAGTTACAGTTGTTAAAGGGCGGGTTGAGGAGCCAACAA 2580
               ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        2521 TGTGTGGTCACTCGCCTCCGCTAGTTACAGTTGTTAAAGGGCGGGTTGAGGAGCCAACAA 2580

Qy        2581 GGTGTGAAAAGCCAGAATGTGCAGCACGGAATGCTATGTCTCTTATTCACAATCGATGCA 2640
               ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        2581 GGTGTGAAAAGCCAGAATGTGCAGCACGGAATGCTATGTCTCTTATTCACAATCGATGCA 2640

Qy        2641 CTTTTGCAAATAAGCAGATAGTGCGTCTTCAAGAAACTCCAGATGCCATTCCTGAAGGAG 2700
               ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        2641 CTTTTGCAAATAAGCAGATAGTGCGTCTTCAAGAAACTCCAGATGCCATTCCTGAAGGAG 2700

Qy        2701 AGACTCCACACACAGTCAGCATGTGTTTATACAACACTATGGTTGATGCTGTGAAGCCTG 2760
               ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        2701 AGACTCCACACACAGTCAGCATGTGTTTATACAACACTATGGTTGATGCTGTGAAGCCTG 2760

Qy        2761 GAGATCGTATTGAGGTAACAGGAGTTTTCAAGGCCATGGCAGTTCGAGTTGGTCCGAATC 2820
               ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        2761 GAGATCGTATTGAGGTAACAGGAGTTTTCAAGGCCATGGCAGTTCGAGTTGGTCCGAATC 2820

Qy        2821 AACGAACATTACGAGCATTGTATAAGACCTACATCGATTGCGTGCACGTCAAGAAGTCTG 2880
               ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        2821 AACGAACATTACGAGCATTGTATAAGACCTACATCGATTGCGTGCACGTCAAGAAGTCTG 2880

Qy        2881 ACAGGGGTCGACTGCAAACTGAAGATCCTATGGAGATGGATAAGGAGAATGATATGTATG 2940
               ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        2881 ACAGGGGTCGACTGCAAACTGAAGATCCTATGGAGATGGATAAGGAGAATGATATGTATG 2940

Qy        2941 CTGGGTATCATGAAAGTGATACTTCAGAAGCTGCTAATGAAGCAAAGATTCAAAAACTTA 3000
               ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        2941 CTGGGTATCATGAAAGTGATACTTCAGAAGCTGCTAATGAAGCAAAGATTCAAAAACTTA 3000

Qy        3001 AAGAGCTGTCCAAGCTCCCGGGCATTTATGATAGACTTTCAAGGTCGCTGGCTCCAAGCA 3060
               ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        3001 AAGAGCTGTCCAAGCTCCCGGGCATTTATGATAGACTTTCAAGGTCGCTGGCTCCAAGCA 3060

Qy        3061 TTTGGGAGCTTGAAGATATTAAAAAGGGTCTTCTTTGCCAGCTCTTTGGTGGGAAGGCTA 3120
               ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        3061 TTTGGGAGCTTGAAGATATTAAAAAGGGTCTTCTTTGCCAGCTCTTTGGTGGGAAGGCTA 3120

Qy        3121 AGAAAATTCCATCTGGAGCATCTTTCCGAGGTGACATCAATGTTTTACTTGTTGGGGACC 3180
               ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        3121 AGAAAATTCCATCTGGAGCATCTTTCCGAGGTGACATCAATGTTTTACTTGTTGGGGACC 3180

Qy        3181 CTGGTACCAGTAAATCTCAGCTGCTTCAGTATGTGCACAAGATAGCTCCTCGTGGAATCT 3240
               ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        3181 CTGGTACCAGTAAATCTCAGCTGCTTCAGTATGTGCACAAGATAGCTCCTCGTGGAATCT 3240

Qy        3241 ACACTAGTGGGCGAGGAAGTTCGGCGGTTGGGCTGACAGCGTATGTAACGAAGGATCCAG 3300
               ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        3241 ACACTAGTGGGCGAGGAAGTTCGGCGGTTGGGCTGACAGCGTATGTAACGAAGGATCCAG 3300

Qy        3301 AAACTCGAGAGACGGTATTGGAGAGCGGAGCTTTGGTTCTTAGTGATCGTGGGATATGCT 3360
               ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        3301 AAACTCGAGAGACGGTATTGGAGAGCGGAGCTTTGGTTCTTAGTGATCGTGGGATATGCT 3360

Qy        3361 GTATCGATGAGTTCGACAAAATGTCTGATAATGCCCGAAGCATGCTTCATGAGGTAATGG 3420
               ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        3361 GTATCGATGAGTTCGACAAAATGTCTGATAATGCCCGAAGCATGCTTCATGAGGTAATGG 3420

Qy        3421 AGCAACAAACGGTATCTGTAGCCAAAGGGGGTATCATTGCCTCGCTGAACGCTCGGACGT 3480
               ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        3421 AGCAACAAACGGTATCTGTAGCCAAAGGGGGTATCATTGCCTCGCTGAACGCTCGGACGT 3480

Qy        3481 CTGTCCTTGCATGTGCAAATCCTAGTGGGTCCCGATACAATGCGCGCCTTTCTGTGATTG 3540
               ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        3481 CTGTCCTTGCATGTGCAAATCCTAGTGGGTCCCGATACAATGCGCGCCTTTCTGTGATTG 3540

Qy        3541 ATAACATCCAGCTTCCTCCAACTCTACTTTCTAGATTTGATTTAATTTACTTAATGCTCG 3600
               ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        3541 ATAACATCCAGCTTCCTCCAACTCTACTTTCTAGATTTGATTTAATTTACTTAATGCTCG 3600

Qy        3601 ACAAACCAGACGAGCAAAACGATCGTCGTCTCGCCAGGCATCTCGTGGCTTTACACTATG 3660
               ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        3601 ACAAACCAGACGAGCAAAACGATCGTCGTCTCGCCAGGCATCTCGTGGCTTTACACTATG 3660

Qy        3661 AAAACTATGAAGTTTCAAAGCAGGACGCCTTAGATCTACAAACACTTACCGCGTATATCA 3720
               ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        3661 AAAACTATGAAGTTTCAAAGCAGGACGCCTTAGATCTACAAACACTTACCGCGTATATCA 3720

Qy        3721 CCTATGCTCGTCAGCATGTACATCCTACATTAAGTGATGAAGCTGCTGAAGATTTGATTA 3780
               ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        3721 CCTATGCTCGTCAGCATGTACATCCTACATTAAGTGATGAAGCTGCTGAAGATTTGATTA 3780

Qy        3781 ATGGCTATGTTGAGATGCGCCAAAAGGGCAACTTTCCTGGAAGCAGTAAAAAGGTGATAA 3840
               ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        3781 ATGGCTATGTTGAGATGCGCCAAAAGGGCAACTTTCCTGGAAGCAGTAAAAAGGTGATAA 3840

Qy        3841 CAGCCACACCTCGGCAACTCGAAAGTATGATTCGTATCAGTGAAGCCCTAGCTCGAATGA 3900
               ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        3841 CAGCCACACCTCGGCAACTCGAAAGTATGATTCGTATCAGTGAAGCCCTAGCTCGAATGA 3900

Qy        3901 GATTTTCTGAAGTGGTAGAGAAAGTTGATGCAGCAGAAGCTGTGCGCCTTTTAGACGTCG 3960
               ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        3901 GATTTTCTGAAGTGGTAGAGAAAGTTGATGCAGCAGAAGCTGTGCGCCTTTTAGACGTCG 3960

Qy        3961 CTTTGCAGCAATCTGCTACTGATCATGCAACAGGTACGATAGACATGGATCTTATCACGA 4020
               ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        3961 CTTTGCAGCAATCTGCTACTGATCATGCAACAGGTACGATAGACATGGATCTTATCACGA 4020

Qy        4021 CTGGAGTGTCGGCCAGCGAGCGTATTCGTCGGGCCAACTTGCTAGCTGCTCTGCGAGAGC 4080
               ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        4021 CTGGAGTGTCGGCCAGCGAGCGTATTCGTCGGGCCAACTTGCTAGCTGCTCTGCGAGAGC 4080

Qy        4081 TTATAGCAGATAAAATTTCACCTGGCAGCTCCTCTGGCTTGAAGACCAGTCAGCTTCTTG 4140
               ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        4081 TTATAGCAGATAAAATTTCACCTGGCAGCTCCTCTGGCTTGAAGACCAGTCAGCTTCTTG 4140

Qy        4141 AGGATATCCGGAGCCAAAGCAGTGTGGACGTTAGTTTGCAGGATATTAAAAAATGCTCTGG 4200
               ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        4141 AGGATATCCGGAGCCAAAGCAGTGTGGACGTTAGTTTGCAGGATATTAAAAAATGCTCTGG 4200

Qy        4201 GTAGCCTCCAAGGAGAAGGCTTTCTTACTGTCCATGGTGACATAGTCAAGAGAGTTTGAG 4260
               ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        4201 GTAGCCTCCAAGGAGAAGGCTTTCTTACTGTCCATGGTGACATAGTCAAGAGAGTTTGAG 4260

Qy        4261 ACAGTTTCTAACTGTTCGAATCCATGAGCTATAACTCTGAACGAAAGGAAAACCTCCAG 4320
               ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        4261 ACAGTTTCTAACTGTTCGAATCCATGAGCTATAACTCTGAACGAAAGGAAAACCTCCAG 4320

Qy        4321 TTTCCCATGCGCAATTCCCAGAGCTCGC 4348
               ||||||||||||||||||||||||||||
Db        4321 TTTCCCATGCGCAATTCCCAGAGCTCGC 4348
```

```
RESULT 2
AX281543
LOCUS       AX281543                4348 bp    DNA     linear   PAT 02-NOV-2001
DEFINITION  Sequence 5 from Patent WO0177354.
ACCESSION   AX281543
VERSION     AX281543.1  GI:16608798
KEYWORDS    .
SOURCE      Physcomitrella patens
  ORGANISM  Physcomitrella patens
            Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Bryophyta;
            Bryopsida; Funariidae; Funariales; Funariaceae; Physcomitrella.
REFERENCE   1
  AUTHORS   costa e Silva,O.D., Bohnert,H.J., van Thielen,N., Chen,R. and
```

```
          Sarria-Millan,R.
  TITLE   Cell cycle stress-related proteins and methods of use in plants
  JOURNAL Patent: WO 0177354-A 5 18-OCT-2001;
          BASF Plant Science GmbH (DE)
FEATURES             Location/Qualifiers
     source          1. .4348
                     /organism="Physcomitrella patens"
                     /mol_type="unassigned DNA"
                     /db_xref="taxon:3218"
ORIGIN

  Query Match             100.0%;  Score 4348;  DB 6;  Length 4348;
  Best Local Similarity   100.0%;  Pred. No. 0;
  Matches 4348;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy          1 ATGGCGCGCCGCACTCACGTGAGGAATTGCACCTCCTTGTTCTGCGACGGTTCCATTCTT 60
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db          1 ATGGCGCGCCGCACTCACGTGAGGAATTGCACCTCCTTGTTCTGCGACGGTTCCATTCTT 60

Qy         61 TTTGGTTTTTAGTTTGCAAATCTTGATCGTGGAGTTGAGAAAAAGGGCGGTTCGTTGTCT 120
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db         61 TTTGGTTTTTAGTTTGCAAATCTTGATCGTGGAGTTGAGAAAAAGGGCGGTTCGTTGTCT 120

Qy        121 TGAGGTGTTCTTGTTGATTGTTCGTCATGGAAAATAATGATGCACTTGACATTGGAGCCG 180
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        121 TGAGGTGTTCTTGTTGATTGTTCGTCATGGAAAATAATGATGCACTTGACATTGGAGCCG 180

Qy        181 TGTCGTCCCCATATCCTTCGCAATCTGAAGGAGTGTCTACGCCATTGCCGCAAGTAACAT 240
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        181 TGTCGTCCCCATATCCTTCGCAATCTGAAGGAGTGTCTACGCCATTGCCGCAAGTAACAT 240

Qy        241 CACCGAGCTTCGACAATGCAGCCTCACCCGTGGCCGGGCGGAGGGCCGTACGGCAGACCC 300
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        241 CACCGAGCTTCGACAATGCAGCCTCACCCGTGGCCGGGCGGAGGGCCGTACGGCAGACCC 300

Qy        301 CTACATCTGCAGTTCGAAGGAGAGGGAGAGAAACGGATTCCGCTCGTCGTAGGAGGAGTC 360
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        301 CTACATCTGCAGTTCGAAGGAGAGGGAGAGAAACGGATTCCGCTCGTCGTAGGAGGAGTC 360

Qy        361 GATCTCGCAGTTTAGGCAATTCTGTTTATAGTTCCCCTTACGATGCGGGGACTCCTGGAA 420
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        361 GATCTCGCAGTTTAGGCAATTCTGTTTATAGTTCCCCTTACGATGCGGGGACTCCTGGAA 420

Qy        421 CTCCTGGAACTCCAGTGGCTACTCCGGTTTACGCTACCCCAGTCGGTACACCTATGGGTA 480
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        421 CTCCTGGAACTCCAGTGGCTACTCCGGTTTACGCTACCCCAGTCGGTACACCTATGGGTA 480

Qy        481 CCCCATCGTTCCATCGTGGCACGCCACAGTACAAACAGCGCAGTGAGCTTGGTTCCCAGG 540
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        481 CCCCATCGTTCCATCGTGGCACGCCACAGTACAAACAGCGCAGTGAGCTTGGTTCCCAGG 540

Qy        541 GGAAGCCTCTACATCGGAGACGTCGATCTCAATCCAGAGAACCCGGGCATCGATCTCCTT 600
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        541 GGAAGCCTCTACATCGGAGACGTCGATCTCAATCCAGAGAACCCGGGCATCGATCTCCTT 600

Qy        601 CAAGGGAACCTAGTGCTGATGGGCGTCCCTCTGAATCTGCTGAGCCAGATGACACTTTGG 660
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        601 CAAGGGAACCTAGTGCTGATGGGCGTCCCTCTGAATCTGCTGAGCCAGATGACACTTTGG 660

Qy        661 GTGGAGAATATGCTTATGTTTGGGGGACGAATGTTAACATTCCAGATGTGCTTAGGGCGA 720
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        661 GTGGAGAATATGCTTATGTTTGGGGGACGAATGTTAACATTCCAGATGTGCTTAGGGCGA 720

Qy        721 TTCGTCGATTTCTCCACAATTATCGTTCGAGTGCTCATGATCTTAATTCCAAGTACATCC 780
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        721 TTCGTCGATTTCTCCACAATTATCGTTCGAGTGCTCATGATCTTAATTCCAAGTACATCC 780

Qy        781 AGATCATAGAGGAGACTGTGGAGCGTGAGGAGGATACTCTAAATATCGACATGTCAGACA 840
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        781 AGATCATAGAGGAGACTGTGGAGCGTGAGGAGGATACTCTAAATATCGACATGTCAGACA 840

Qy        841 TTTATGACCATGATCCTGATCTATACGCAAAAATTGTTCGATACCCACTCGACATCATCC 900
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        841 TTTATGACCATGATCCTGATCTATACGCAAAAATTGTTCGATACCCACTCGACATCATCC 900

Qy        901 CCCTGTTGGACACTGAGTGTCAGGAAGTTGCTACCTCTTTACTACCAACGTTTGAGAAGC 960
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        901 CCCTGTTGGACACTGAGTGTCAGGAAGTTGCTACCTCTTTACTACCAACGTTTGAGAAGC 960

Qy        961 ATATTGAGGCCAGACCTTTCAATCTCAAAGCATCGGTGCACATGCGTGAACTCAACCCTT 1020
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        961 ATATTGAGGCCAGACCTTTCAATCTCAAAGCATCGGTGCACATGCGTGAACTCAACCCTT 1020

Qy       1021 CAGATATAGACAAATTGGTTTCTGTTAAAGGAATGGTTATCCGGTGCAGTTCTATCATAC 1080
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1021 CAGATATAGACAAATTGGTTTCTGTTAAAGGAATGGTTATCCGGTGCAGTTCTATCATAC 1080

Qy       1081 CTGAAATTAAGGGGGCCTTCTTCAAATGTTTAGTGTGTGGTCACTCGCCTCCGCTAGTTA 1140
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1081 CTGAAATTAAGGGGGCCTTCTTCAAATGTTTAGTGTGTGGTCACTCGCCTCCGCTAGTTA 1140

Qy       1141 CAGTTGTTAAAGGGCGGGTTGAGGAGCCAACAAGGTGTGAAAAGCCAGAATGTGCAGCAC 1200
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1141 CAGTTGTTAAAGGGCGGGTTGAGGAGCCAACAAGGTGTGAAAAGCCAGAATGTGCAGCAC 1200

Qy       1201 GGAATGCTATGTCTCTTATTCACAATCGATGCACTTTTGCAAATAAGCAGATAGTGCGTC 1260
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1201 GGAATGCTATGTCTCTTATTCACAATCGATGCACTTTTGCAAATAAGCAGATAGTGCGTC 1260

Qy       1261 TTCAAGAAACTCCAGATGCCATTCCTGAAGGAGAGACTCCACACACAGTCAGCATGTGTT 1320
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1261 TTCAAGAAACTCCAGATGCCATTCCTGAAGGAGAGACTCCACACACAGTCAGCATGTGTT 1320

Qy       1321 TATACAACACTATGGTTGATGCTGTGAAGCCTGGAGATCGTATTGAGGTAACAGGAGTTT 1380
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1321 TATACAACACTATGGTTGATGCTGTGAAGCCTGGAGATCGTATTGAGGTAACAGGAGTTT 1380

Qy       1381 TCAAGGCCATGGCAGTTCGAGTTGGTCATGGCGCGCCGCACTCACGTGAGGAATTGCACC 1440
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1381 TCAAGGCCATGGCAGTTCGAGTTGGTCATGGCGCGCCGCACTCACGTGAGGAATTGCACC 1440

Qy       1441 TCCTTGTTCTGCGACGGTTCCATTCTTTTTGGTTTTTAGTTTGCAAATCTTGATCGTGGA 1500
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1441 TCCTTGTTCTGCGACGGTTCCATTCTTTTTGGTTTTTAGTTTGCAAATCTTGATCGTGGA 1500

Qy       1501 GTTGAGAAAAAGGGCGGTTCGTTGTCTTGAGGTGTTCTTGTTGATTGTTCGTCATGGAAA 1560
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1501 GTTGAGAAAAAGGGCGGTTCGTTGTCTTGAGGTGTTCTTGTTGATTGTTCGTCATGGAAA 1560

Qy       1561 ATAATGATGCACTTGACATTGGAGCCGTGTCGTCCCCATATCCTTCGCAATCTGAAGGAG 1620
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1561 ATAATGATGCACTTGACATTGGAGCCGTGTCGTCCCCATATCCTTCGCAATCTGAAGGAG 1620

Qy       1621 TGTCTACGCCATTGCCGCAAGTAACATCACCGAGCTTCGACAATGCAGCCTCACCCGTGG 1680
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1621 TGTCTACGCCATTGCCGCAAGTAACATCACCGAGCTTCGACAATGCAGCCTCACCCGTGG 1680

Qy       1681 CCGGGCGGAGGGCCGTACGGCAGACCCCTACATCTGCAGTTCGAAGGAGAGGGAGAGAAA 1740
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1681 CCGGGCGGAGGGCCGTACGGCAGACCCCTACATCTGCAGTTCGAAGGAGAGGGAGAGAAA 1740

Qy       1741 CGGATTCCGCTCGTCGTAGGAGGAGTCGATCTCGCAGTTTAGGCAATTCTGTTTATAGTT 1800
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1741 CGGATTCCGCTCGTCGTAGGAGGAGTCGATCTCGCAGTTTAGGCAATTCTGTTTATAGTT 1800

Qy       1801 CCCCTTACGATGCGGGGACTCCTGGAACTCCTGGAACTCCAGTGGCTACTCCGGTTTACG 1860
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1801 CCCCTTACGATGCGGGGACTCCTGGAACTCCTGGAACTCCAGTGGCTACTCCGGTTTACG 1860

Qy       1861 CTACCCCAGTCGGTACACCTATGGGTACCCCATCGTTCCATCGTGGCACGCCACAGTACA 1920
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1861 CTACCCCAGTCGGTACACCTATGGGTACCCCATCGTTCCATCGTGGCACGCCACAGTACA 1920

Qy       1921 AACAGCGCAGTGAGCTTGGTTCCCAGGGGAAGCCTCTACATCGGAGACGTCGATCTCAAT 1980
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1921 AACAGCGCAGTGAGCTTGGTTCCCAGGGGAAGCCTCTACATCGGAGACGTCGATCTCAAT 1980
```

```
Qy      1981 CCAGAGAACCCGGGCATCGATCTCCTTCAAGGGAACCTAGTGCTGATGGGCGTCCCTCTG 2040
             ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      1981 CCAGAGAACCCGGGCATCGATCTCCTTCAAGGGAACCTAGTGCTGATGGGCGTCCCTCTG 2040

Qy      2041 AATCTGCTGAGCCAGATGACACTTTGGGTGGAGAATATGCTTATGTTTGGGGGACGAATG 2100
             ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      2041 AATCTGCTGAGCCAGATGACACTTTGGGTGGAGAATATGCTTATGTTTGGGGGACGAATG 2100

Qy      2101 TTAACATTCCAGATGTGCTTAGGGCGATTCGTCGATTTCTCCACAATTATCGTTCGAGTG 2160
             ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      2101 TTAACATTCCAGATGTGCTTAGGGCGATTCGTCGATTTCTCCACAATTATCGTTCGAGTG 2160

Qy      2161 CTCATGATCTTAATTCCAAGTACATCCAGATCATAGAGGAGACTGTGGAGCGTGAGGAGG 2220
             ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      2161 CTCATGATCTTAATTCCAAGTACATCCAGATCATAGAGGAGACTGTGGAGCGTGAGGAGG 2220

Qy      2221 ATACTCTAAATATCGACATGTCAGACATTTATGACCATGATCCTGATCTATACGCAAAAA 2280
             ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      2221 ATACTCTAAATATCGACATGTCAGACATTTATGACCATGATCCTGATCTATACGCAAAAA 2280

Qy      2281 TTGTTCGATACCCACTCGACATCATCCCCCTGTTGGACACTGAGTGTCAGGAAGTTGCTA 2340
             ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      2281 TTGTTCGATACCCACTCGACATCATCCCCCTGTTGGACACTGAGTGTCAGGAAGTTGCTA 2340

Qy      2341 CCTCTTTACTACCAACGTTTGAGAAGCATATTGAGGCCAGACCTTTCAATCTCAAAGCAT 2400
             ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      2341 CCTCTTTACTACCAACGTTTGAGAAGCATATTGAGGCCAGACCTTTCAATCTCAAAGCAT 2400

Qy      2401 CGGTGCACATGCGTGAACTCAACCCTTCAGATATAGACAAATTGGTTTCTGTTAAAGGAA 2460
             ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Bb      2401 CGGTGCACATGCGTGAACTCAACCCTTCAGATATAGACAAATTGGTTTCTGTTAAAGGAA 2460

Qy      2461 TGGTTATCCGGTGCAGTTCTATCATACCTGAAATTAAGGGGGCCTTCTTCAAATGTTTAG 2520
             ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      2461 TGGTTATCCGGTGCAGTTCTATCATACCTGAAATTAAGGGGGCCTTCTTCAAATGTTTAG 2520

Qy      2521 TGTGTGGTCACTCGCCTCCGCTAGTTACAGTTGTTAAAGGGCGGGTTGAGGAGCCAACAA 2580
             ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      2521 TGTGTGGTCACTCGCCTCCGCTAGTTACAGTTGTTAAAGGGCGGGTTGAGGAGCCAACAA 2580

Qy      2581 GGTGTGAAAAGCCAGAATGTGCAGCACGGAATGCTATGTCTCTTATTCACAATCGATGCA 2640
             ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      2581 GGTGTGAAAAGCCAGAATGTGCAGCACGGAATGCTATGTCTCTTATTCACAATCGATGCA 2640

Qy      2641 CTTTTGCAAATAAGCAGATAGTGCGTCTTCAAGAAACTCCAGATGCCATTCCTGAAGGAG 2700
             ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      2641 CTTTTGCAAATAAGCAGATAGTGCGTCTTCAAGAAACTCCAGATGCCATTCCTGAAGGAG 2700

Qy      2701 AGACTCCACACACAGTCAGCATGTGTTTATACAACACTATGGTTGATGCTGTGAAGCCTG 2760
             ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      2701 AGACTCCACACACAGTCAGCATGTGTTTATACAACACTATGGTTGATGCTGTGAAGCCTG 2760

Qy      2761 GAGATCGTATTGAGGTAACAGGAGTTTTCAAGGCCATGGCAGTTCGAGTTGGTCCGAATC 2820
             ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      2761 GAGATCGTATTGAGGTAACAGGAGTTTTCAAGGCCATGGCAGTTCGAGTTGGTCCGAATC 2820

Qy      2821 AACGAACATTACGAGCATTGTATAAGACCTACATCGATTGCGTGCACGTCAAGAAGTCTG 2880
             ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      2821 AACGAACATTACGAGCATTGTATAAGACCTACATCGATTGCGTGCACGTCAAGAAGTCTG 2880

Qy      2881 ACAGGGGTCGACTGCAAACTGAAGATCCTATGGAGATGGATAAGGAGAATGATATGTATG 2940
             ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      2881 ACAGGGGTCGACTGCAAACTGAAGATCCTATGGAGATGGATAAGGAGAATGATATGTATG 2940

Qy      2941 CTGGGTATCATGAAAGTGATACTTCAGAAGCTGCTAATGAAGCAAAGATTCAAAAACTTA 3000
             ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      2941 CTGGGTATCATGAAAGTGATACTTCAGAAGCTGCTAATGAAGCAAAGATTCAAAAACTTA 3000

Qy      3001 AAGAGCTGTCCAAGCTCCCGGGCATTTATGATAGACTTTCAAGGTCGCTGGCTCCAAGCA 3060
             ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      3001 AAGAGCTGTCCAAGCTCCCGGGCATTTATGATAGACTTTCAAGGTCGCTGGCTCCAAGCA 3060

Qy      3061 TTTGGGAGCTTGAAGATATTAAAAAGGGTCTTCTTTGCCAGCTCTTTGGTGGAAGGCTA 3120
             ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      3061 TTTGGGAGCTTGAAGATATTAAAAAGGGTCTTCTTTGCCAGCTCTTTGGTGGAAGGCTA 3120

Qy      3121 AGAAAATTCCATCTGGAGCATCTTTCCGAGGTGACATCAATGTTTTACTTGTTGGGGACC 3180
             ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      3121 AGAAAATTCCATCTGGAGCATCTTTCCGAGGTGACATCAATGTTTTACTTGTTGGGGACC 3180

Qy      3181 CTGGTACCAGTAAATCTCAGCTGCTTCAGTATGTGCACAAGATAGCTCCTCGTGGAATCT 3240
             ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      3181 CTGGTACCAGTAAATCTCAGCTGCTTCAGTATGTGCACAAGATAGCTCCTCGTGGAATCT 3240

Qy      3241 ACACTAGTGGGCGAGGAAGTTCGGCGGTTGGGCTGACAGCGTATGTAACGAAGGATCCAG 3300
             ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      3241 ACACTAGTGGGCGAGGAAGTTCGGCGGTTGGGCTGACAGCGTATGTAACGAAGGATCCAG 3300

Qy      3301 AAACTCGAGAGACGGTATTGGAGAGCGGAGCTTTGGTTCTTAGTGATCGTGGGATATGCT 3360
             ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      3301 AAACTCGAGAGACGGTATTGGAGAGCGGAGCTTTGGTTCTTAGTGATCGTGGGATATGCT 3360

Qy      3361 GTATCGATGAGTTCGACAAAATGTCTGATAATGCCCGAAGCATGCTTCATGAGGTAATGG 3420
             ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      3361 GTATCGATGAGTTCGACAAAATGTCTGATAATGCCCGAAGCATGCTTCATGAGGTAATGG 3420

Qy      3421 AGCAACAAACGGTATCTGTAGCCAAAGGGGGTATCATTGCCTCGCTGAACGCTCGGACGT 3480
             ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      3421 AGCAACAAACGGTATCTGTAGCCAAAGGGGGTATCATTGCCTCGCTGAACGCTCGGACGT 3480

Qy      3481 CTGTCCTTGCATGTGCAAATCCTAGTGGGTCCCGATACAATGCGCGCCTTTCTGTGATTG 3540
             ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      3481 CTGTCCTTGCATGTGCAAATCCTAGTGGGTCCCGATACAATGCGCGCCTTTCTGTGATTG 3540

Qy      3541 ATAACATCCAGCTTCCTCCAACTCTACTTTCTAGATTTGATTTAATTTACTTAATGCTCG 3600
             ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      3541 ATAACATCCAGCTTCCTCCAACTCTACTTTCTAGATTTGATTTAATTTACTTAATGCTCG 3600

Qy      3601 ACAAACCAGACGAGCAAAACGATCGTCGTCTCGCCAGGCATCTCGTGGCTTTACACTATG 3660
             ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      3601 ACAAACCAGACGAGCAAAACGATCGTCGTCTCGCCAGGCATCTCGTGGCTTTACACTATG 3660

Qy      3661 AAAACTATGAAGTTTCAAAGCAGGACGCCTTAGATCTACAAACACTTACCGCGTATATCA 3720
             ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      3661 AAAACTATGAAGTTTCAAAGCAGGACGCCTTAGATCTACAAACACTTACCGCGTATATCA 3720

Qy      3721 CCTATGCTCGTCAGCATGTACATCCTACATTAAGTGATGAAGCTGCTGAAGATTTGATTA 3780
             ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      3721 CCTATGCTCGTCAGCATGTACATCCTACATTAAGTGATGAAGCTGCTGAAGATTTGATTA 3780

Qy      3781 ATGGCTATGTTGAGATGCGCCAAAAGGGCAACTTTCCTGGAAGCAGTAAAAAGGTGATAA 3840
             ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      3781 ATGGCTATGTTGAGATGCGCCAAAAGGGCAACTTTCCTGGAAGCAGTAAAAAGGTGATAA 3840

Qy      3841 CAGCCACACCTCGGCAACTCGAAAGTATGATTCGTATCAGTGAAGCCCTAGCTCGAATGA 3900
             ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      3841 CAGCCACACCTCGGCAACTCGAAAGTATGATTCGTATCAGTGAAGCCCTAGCTCGAATGA 3900

Qy      3901 GATTTTCTGAAGTGGTAGAGAAAGTTGATGCAGCAGAAGCTGTGCGCCTTTTAGACGTCG 3960
             ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      3901 GATTTTCTGAAGTGGTAGAGAAAGTTGATGCAGCAGAAGCTGTGCGCCTTTTAGACGTCG 3960

Qy      3961 CTTTGCAGCAATCTGCTACTGATCATGCAACAGGTACGATAGACATGGATCTTATCACGA 4020
             ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      3961 CTTTGCAGCAATCTGCTACTGATCATGCAACAGGTACGATAGACATGGATCTTATCACGA 4020

Qy      4021 CTGGAGTGTCGGCCAGCGAGCGTATTCGTCGGGCCAACTTGCTAGCTGCTCTGCGAGAGC 4080
             ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      4021 CTGGAGTGTCGGCCAGCGAGCGTATTCGTCGGGCCAACTTGCTAGCTGCTCTGCGAGAGC 4080

Qy      4081 TTATAGCAGATAAAATTTCACCTGGCAGCTCCTCTGGCTTGAAGACCAGTCAGCTTCTTG 4140
             ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      4081 TTATAGCAGATAAAATTTCACCTGGCAGCTCCTCTGGCTTGAAGACCAGTCAGCTTCTTG 4140

Qy      4141 AGGATATCCGGAGCCAAAGCAGTGTGGACGTTAGTTTGCAGGATATTAAAAATGCTCTGG 4200
```

```
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       4141 AGGATATCCGGAGCCAAAGCAGTGTGGACGTTAGTTTGCAGGATATTAAAAATGCTCTGG 4200

Qy       4201 GTAGCCTCCAAGGAGAAGGCTTTCTTACTGTCCATGGTGACATAGTCAAGAGAGTTTGAG 4260
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       4201 GTAGCCTCCAAGGAGAAGGCTTTCTTACTGTCCATGGTGACATAGTCAAGAGAGTTTGAG 4260

Qy       4261 ACAGTTTCTAACTGTTCGAATCCATGAGCTATAACTCTGAACGAAAGGGAAAACCTCCAG 4320
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       4261 ACAGTTTCTAACTGTTCGAATCCATGAGCTATAACTCTGAACGAAAGGGAAAACCTCCAG 4320

Qy       4321 TTTCCCATGCGCAATTCCCAGAGCTCGC 4348
              ||||||||||||||||||||||||||||
Db       4321 TTTCCCATGCGCAATTCCCAGAGCTCGC 4348


RESULT 3
AR490065
LOCUS       AR490065                 723 bp    DNA     linear   PAT 15-MAY-2004
DEFINITION  Sequence 2 from patent US 6710229.
ACCESSION   AR490065
VERSION     AR490065.1  GI:47257188
KEYWORDS    .
SOURCE      Unknown.
  ORGANISM  Unknown.
            Unclassified.
REFERENCE   1  (bases 1 to 723)
  AUTHORS   da Costa e Silva,O., Bohnert,H.J., van Thielen,N., Chen,R. and
            Sarria-Millan,R.
  TITLE     Cell cycle stress-related proteins and methods of use in plants
  JOURNAL   Patent: US 6710229-A 2 23-MAR-2004;
            BASF Plant Science GmbH; Ludwigshafen;
            WOX;
FEATURES             Location/Qualifiers
     source          1..723
                     /organism="unknown"
                     /mol_type="genomic DNA"
ORIGIN

  Query Match             15.5%;  Score 673;  DB 6;  Length 723;
  Best Local Similarity   97.1%;  Pred. No. 2.5e-189;
  Matches  696;  Conservative    0;  Mismatches   20;  Indels    1;  Gaps    1;

Qy       2753 GAAGCCTGGAGATCGTATTGAGGTAACAGGAGTTTTCAAGGCCATGGCAGTTCGAGTTGG 2812
              | | |  |||||||||||||||||||||||||||||||||||||||||||||||||||||
Db          3 GCACCAGGGAGATCGTATTGAGGTAACAGGAGTTTTCAAGGCCATGGCAGTTCGAGTTGG 62

Qy       2813 TCCGAATCAACGAACATTACGAGCATTGTATAAGACCTACATCGATTGCGTGCACGTCAA 2872
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db         63 TCCGAATCAACGAACATTACGAGCATTGTATAAGACCTACATCGATTGCGTGCACGTCAA 122

Qy       2873 GAAGTCTGACAGGGGTCGACTGCAAACTGAAGATCCTATGGAGATGGATAAGGAGAATGA 2932
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        123 GAAGTCTGACAGGGGTCGACTGCAAACTGAAGATCCTATGGAGATGGATAAGGAGAATGA 182

Qy       2933 TATGTATGCTGGGTATCATGAAAGTGATACTTCAGAAGCTGCTAATGAAGCAAAGATTCA 2992
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        183 TATGTATGCTGGGTATCATGAAAGTGATACTTCAGAAGCTGCTAATGAAGCAAAGATTCA 242

Qy       2993 AAAACTTAAAGAGCTGTCCAAGCTCCCGGGCATTTATGATAGACTTTCAAGGTCGCTGGC 3052
              ||||||||||||||||||||||||||||| ||||||||||||||||||||||||||||||
Db        243 AAAACTTAAAGAGCTGTCCAAGCTCCCGGACATTTATGATAGACTTTCAAGGTCGCTGGC 302

Qy       3053 TCCAAGCATTTGGGAGCTTGAAGATATTAAAAAGGGTCTTCTTTGCCAGCTCTTTGGTGG 3112
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        303 TCCAAGCATTTGGGAGCTTGAAGATATTAAAAAGGGTCTTCTTTGCCAGCTCTTTGGTGG 362

Qy       3113 GAAGGCTAAGAAAATTCCATCTGGAGCATCTTTCCGAGGTGACATCAATGTTTTACTTGT 3172
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        363 GAAGGCTAAGAAAATTCCATCTGGAGCATCTTTCCGAGGTGACATCAATGTTTTACTTGT 422

Qy       3173 TGGGGACCCTGGTACCAGTAAATCTCAGCTGCTTCAGTATGTGCACAAGATAGCTCCTCG 3232
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        423 TGGGGACCCTGGTACCAGTAAATCTCAGCTGCTTCAGTATGTGCACAAGATAGCTCCTCG 482

Qy       3233 TGGAATCTACACTAGTGGGCGAGGAAGTTCGGCGGTTGGGCTGACAGCGTATGT-AACGA 3291
              |||||||||||||||||||||||||||||||||||||||||||||||||||||| |||||
Db        483 TGGAATCTACACTAGTGGGCGAGGAAGTTCGGCGGTTGGGCTGACAGCGTATGTAAACGA 542

Qy       3292 AGGATCCAGAAACTCGAGAGACGGTATTGGAGAGCGGAGCTTTGGTTCTTAGTGATCGTG 3351
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        543 AGGATCCAGAAACTCGAGAGACGGTATTGGAGAGCGGAGCTTTGGTTCTTAGTGATCGTG 602

Qy       3352 GGATATGCTGTATCGATGAGTTCGACAAAATGTCTGATAATGCCCGAAGCATGCTTCATG 3411
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        603 GGATATGCTGTATCGATGAGTTCGACAAAATGTCTGATAATGCCCGAAGCATGCTTCATG 662

Qy       3412 AGGTAATGGAGCAACAAACGGTATCTGTAGCCAAAGGGGGTATCATTGCCTCGCTGA 3468
              ||||||||||||||||||||||||||| | ||||  ||   ||  | |    || ||
Db        663 AGGTAATGGAGCAACAAACGGTATCTGGACCCAAGCGGTTCATGCTCGTGAAGCCGA 719


RESULT 4
AX281540
LOCUS       AX281540                 723 bp    DNA     linear   PAT 02-NOV-2001
DEFINITION  Sequence 2 from Patent WO0177354.
ACCESSION   AX281540
VERSION     AX281540.1  GI:16608795
KEYWORDS    .
SOURCE      Physcomitrella patens
  ORGANISM  Physcomitrella patens
            Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Bryophyta;
            Bryopsida; Funariidae; Funariales; Funariaceae; Physcomitrella.
REFERENCE   1
  AUTHORS   costa e Silva,O.D., Bohnert,H.J., van Thielen,N., Chen,R. and
            Sarria-Millan,R.
  TITLE     Cell cycle stress-related proteins and methods of use in plants
  JOURNAL   Patent: WO 0177354-A 2 18-OCT-2001;
            BASF Plant Science GmbH (DE)
FEATURES             Location/Qualifiers
     source          1..723
                     /organism="Physcomitrella patens"
                     /mol_type="unassigned DNA"
                     /db_xref="taxon:3218"
ORIGIN

  Query Match             15.5%;  Score 673;  DB 6;  Length 723;
  Best Local Similarity   97.1%;  Pred. No. 2.5e-189;
  Matches  696;  Conservative    0;  Mismatches   20;  Indels    1;  Gaps    1;

Qy       2753 GAAGCCTGGAGATCGTATTGAGGTAACAGGAGTTTTCAAGGCCATGGCAGTTCGAGTTGG 2812
              | | |  |||||||||||||||||||||||||||||||||||||||||||||||||||||
Db          3 GCACCAGGGAGATCGTATTGAGGTAACAGGAGTTTTCAAGGCCATGGCAGTTCGAGTTGG 62

Qy       2813 TCCGAATCAACGAACATTACGAGCATTGTATAAGACCTACATCGATTGCGTGCACGTCAA 2872
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db         63 TCCGAATCAACGAACATTACGAGCATTGTATAAGACCTACATCGATTGCGTGCACGTCAA 122

Qy       2873 GAAGTCTGACAGGGGTCGACTGCAAACTGAAGATCCTATGGAGATGGATAAGGAGAATGA 2932
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        123 GAAGTCTGACAGGGGTCGACTGCAAACTGAAGATCCTATGGAGATGGATAAGGAGAATGA 182

Qy       2933 TATGTATGCTGGGTATCATGAAAGTGATACTTCAGAAGCTGCTAATGAAGCAAAGATTCA 2992
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        183 TATGTATGCTGGGTATCATGAAAGTGATACTTCAGAAGCTGCTAATGAAGCAAAGATTCA 242

Qy       2993 AAAACTTAAAGAGCTGTCCAAGCTCCCGGGCATTTATGATAGACTTTCAAGGTCGCTGGC 3052
              ||||||||||||||||||||||||||||| ||||||||||||||||||||||||||||||
Db        243 AAAACTTAAAGAGCTGTCCAAGCTCCCGGACATTTATGATAGACTTTCAAGGTCGCTGGC 302

Qy       3053 TCCAAGCATTTGGGAGCTTGAAGATATTAAAAAGGGTCTTCTTTGCCAGCTCTTTGGTGG 3112
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        303 TCCAAGCATTTGGGAGCTTGAAGATATTAAAAAGGGTCTTCTTTGCCAGCTCTTTGGTGG 362

Qy       3113 GAAGGCTAAGAAAATTCCATCTGGAGCATCTTTCCGAGGTGACATCAATGTTTTACTTGT 3172
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
```

```
Db        363 GAAGGCTAAGAAAATTCCATCTGGAGCATCTTTCCGAGGTGACATCAATGTTTTACTTGT 422

Qy       3173 TGGGGACCCTGGTACCAGTAAATCTCAGCTGCTTCAGTATGTGCACAAGATAGCTCCTCG 3232

Db        423 TGGGGACCCTGGTACCAGTAAATCTCAGCTGCTTCAGTATGTGCACAAGATAGCTCCTCG 482

Qy       3233 TGGAATCTACACTAGTGGGCGAGGAAGTTCGGCGGTTGGGCTGACAGCGTATGT-AACGA 3291

Db        483 TGGAATCTACACTAGTGGGCGAGGAAGTTCGGCGGTTGGGCTGACAGCGTATGTAAACGA 542

Qy       3292 AGGATCCAGAAACTCGAGAGACGGTATTGGAGAGCGGAGCTTTGGTTCTTAGTGATCGTG 3351

Db        543 AGGATCCAGAAACTCGAGAGACGGTATTGGAGAGCGGAGCTTTGGTTCTTAGTGATCGTG 602

Qy       3352 GGATATGCTGTATCGATGAGTTCGACAAAATGTCTGATAATGCCCGAAGCATGCTTCATG 3411

Db        603 GGATATGCTGTATCGATGAGTTCGACAAAATGTCTGATAATGCCCGAAGCATGCTTCATG 662

Qy       3412 AGGTAATGGAGCAACAAACGGTATCTGTAGCCAAAGGGGGTATCATTGCCTCGCTGA 3468

Db        663 AGGTAATGGAGCAACAAACGGTATCTGGACCCAAGCGGTTCATGCTCGTGAAGCCGA 719
```

```
RESULT 5
SPCDC21
LOCUS       SPCDC21                 3667 bp    DNA     linear   PLN 02-MAR-2000
DEFINITION  Schizosaccharomyces pombe cdc21 gene for protein required for S
            phase execution.
ACCESSION   X58824
VERSION     X58824.1  GI:4165292
KEYWORDS    cdc21+ gene; cell division control; mcm4.
SOURCE      Schizosaccharomyces pombe (fission yeast)
  ORGANISM  Schizosaccharomyces pombe
            Eukaryota; Fungi; Ascomycota; Schizosaccharomycetes;
            Schizosaccharomycetales; Schizosaccharomycetaceae;
            Schizosaccharomyces.
REFERENCE   1
  AUTHORS   Coxon,A., Maundrell,K. and Kearsey,S.E.
  TITLE     Fission yeast cdc21+ belongs to a family of proteins involved in an
            early step of chromosome replication
  JOURNAL   Nucleic Acids Res. 20 (21), 5571-5577 (1992)
   PUBMED   1454522
REFERENCE   2
  AUTHORS   Kearsey,S.E.
  TITLE     Direct Submission
  JOURNAL   Submitted (05-APR-1991) S.E. Kearsey, Oxford University, South
            Parks Road, Oxford, UK
  REMARK    revised by [3]
REFERENCE   3  (bases 1 to 3667)
  AUTHORS   Kearsey,S.E.
  TITLE     Direct Submission
  JOURNAL   Submitted (15-JAN-1999) S.E. Kearsey, Oxford University, South
            Parks Road, Oxford, UK
COMMENT     On Jan 21, 1999 this sequence version replaced gi:4926.
FEATURES             Location/Qualifiers
     source          1..3667
                     /organism="Schizosaccharomyces pombe"
                     /mol_type="genomic DNA"
                     /db_xref="taxon:4896"
     gene            641..3436
                     /gene="cdc21"
     CDS             641..3436
                     /gene="cdc21"
                     /codon_start=1
                     /product="cdc21 protein"
                     /protein_id="CAA41628.1"
                     /db_xref="GI:4165293"
                     /db_xref="GOA:P29458"
                     /db_xref="UniProt/Swiss-Prot:P29458"
                     /translation="MSSSQQSGRANELRTPGRANSSSREAVDSSPLFFPASSPGSTRL
                     TTPRTTARTPLASSPLLFESSSPGPNIPQSSRSHLLSQRNDLFLDSSSQRTPRSTRRG
                     DIHSSVQMSTPSRRREVDPQRPGVSTPSSLLFSGSDALTFSQAHPSSEVADDTVRVIW
                     GTNVSIQESIASFRGFLRGFKKKYRPEYRNELMPPPDAEQLVYIEALRNMRIMGLEIL
                     NLDVQDLKHYPPTKKLYHQLYSYPQEIIPIMDQTIKDVMLDLLGTNPPEDVLNDIELK
                     IYKIRPFNLEKCINMRDLNPGDIDKLISIKGLVLRCTPVIPDMKQAFFRCSVCGHCVT
                     VEIDRGRIAEPIKCPREVCGATNAMQLIHNRSEFADKQVIKLQETPDVVPDGQTPHSV
                     SLCVYDELVDSARAGDRIEVTGIFRCVPVRLNPRMRTVKSLFKTYVDVVHIKKQDKRR
                     LGTDPSTLESDIAEDAALQIDEVRKISDEEVEKIQQVSKRDDIYDILSRSLAPSIYEM
                     DDVKKGLLLQLFGGTNKSFHKGASPRYRGDINILMCGDPSTSKSQILKYVHKIAPRGV
                     YTSGKGSSAVGLTAYITRDQDTKQLVLESGALVLSDGGICCIDEFDKMSDATRSILHE
                     VMEQQTVTVAKAGIITTLNARTSILASANPIGSKYNPDLPVTKNIDLPPTLLSRFDLV
                     YLILDRVDETLDRKLANHIVSMYMEDTPEHATDMEVFSVEFLTSYITYARNNINPVIS
                     EEAAKELVNAYVGMRKLGEDVRASEKRITATTRQLESMIRLSEAHAKMHLRNVVEVGD
                     VLEAARLIKTAIKDYATDPATGKISLDLIYVNERETLVPEDMVKELANLISNLTVGGK
                     TMLVSQLLTRFREQSSTRLDASDFEACLGAFREQSSTRLDASDFEACLGALERRGRIK
                     VITSAGHRIVRSIAQTD"
ORIGIN

  Query Match             12.0%;  Score 520.8;  DB 15;  Length 3667;
  Best Local Similarity   59.5%;  Pred. No. 1.2e-143;
  Matches  980;  Conservative    0;  Mismatches  647;  Indels   21;  Gaps    5;
```

```
Qy       2380 GACCTTTCAATCTCAAAGCATCGGTGCACATGCGTGAACTCAACCCTTCAGATATAGACA 2439

Db       1482 GTCCTTTTAATCTTGAAAAATGCATTAATATGCGTGATCTTAATCCTGGTGACATTGACA 1541

Qy       2440 AATTGGTTTCTGTTAAAGGAATGGTTATCCGGTGCAGTTCTATCATACCTGAAATTAAGG 2499

Db       1542 AATTGATCTCTATTAAAGGTCTGGTACTTAGGTGTACACCTGTAATACCCGACATGAAGC 1601

Qy       2500 GGGCCTTCTTCAAATGTTTAGTGTGTGGTCACTCGCCTCCGCTAGTTACAGTTGTTAAAG 2559

Db       1602 AAGCATTTTTTCGATGTTCAGTTTGTGGTCATTGTGTTAC---CGTGGAAATCGATAGAG 1658

Qy       2560 GGCGGGTTGAGGAGCCAACAAGGTGTGAAAAGCCAGAATGTGCAGCACGGAATGCTATGT 2619

Db       1659 GGCGAATTGCTGAGCCCATAAAATGCCCTAGAGAAGTCTGTGGTGCTACAAATGCAATGC 1718

Qy       2620 CTCTTATTCACAATCGATGCACTTTTGCAAATAAGCAGATAGTGCGTCTTCAAGAAACTC 2679

Db       1719 AGCTAATTCATAATCGTTCTGAATTCGCCGATAAGCAAGTTATTAAACTCCAAGAGACAC 1778

Qy       2680 CAGATGCCATTCCTGAAGGAGAGACTCCACACACAGTCAGCATGTGTTTATACAACACTA 2739

Db       1779 CTGACGTGGTACCGGATGGCCAAACCCCACACTCTGTAAGTCTGTGCGTATATGATGAAC 1838

Qy       2740 TGGTTGATGCTGTGAAGCCTGGAGATCGTATTGAGGTAACAGGAGTTTTCAAGGCCATGG 2799

Db       1839 TGGTTGACTCTGCTAGAGCAGGAGATCGTATAGAAGTCACTGGGATTTTTCGCTGTGTAC 1898

Qy       2800 CAGTTCGAGTTGGTCCGAATCAACGAACATTACGAGCATTGTATAAGACCTACATCGATT 2859

Db       1899 CAGTTCGTTTGAACCCTCGCATGCGTACTGTAAAATCATTATTTAAAACTTATGTCGACG 1958

Qy       2860 GCGTGCACGTCAAGAAGTCTGACAGGGGTCGACTGCAAACTGAAGATCCTATGGAGATGG 2919

Db       1959 TTGTTCACATAAAAAAGCAGGACAAACGTCGTTTGGGTAC---GGATCCTTCTACGTTGG 2015

Qy       2920 ATAAGGAGAATGATATGTATGCTGGGTATCATGAAAGTGATACTTCAGAAGCTGCTAATG 2979

Db       2016 AAAGCGATATTGCTGAAGATGCTGCTTTAC---AAATTGATGAAGTACGGAAAATATCAG 2072

Qy       2980 AAGCAAAGATTCAAAAACTTAAAGAGCTGTCCAAGCTCCCGGGCATTTATGATAGACTTT 3039

Db       2073 ACGAAGAAGTGGAAAAAATTCAACAAGTTTCAAAGCGGGACGATATATACGATATTCTTT 2132

Qy       3040 CAAGGTCGCTGGCTCCAAGCATTTGGGAGCTTGAAGATATTAAAAAGGGTCTTCTTTGCC 3099

Db       2133 CCCGGTCTCTGGCACCTAGCATTTACGAAATGGATGATGTTAAAAAGGGGCTTTTGCTAC 2192

Qy       3100 AGCTCTTTGGTGGGAAGGCTAAGAAAATTCCATCTGGAGCATCT------TTCCGAGGTG 3153

Db       2193 AGCTATTTGGTGGAACGAACAAGTCTTTTCATAAAGGCGCAAGTCCGAGATATCGTGGAG 2252

Qy       3154 ACATCAATGTTTTACTTGTTGGGGACCCTGGTACCAGTAAATCTCAGCTGCTTCAGTATG 3213
```

```
Db       2253 ACATAAATATCTTGATGTGTGGTGATCCATCTACGTCTAAGTCTCAAATTTTGAAGTATG 2312

Qy       3214 TGCACAAGATAGCTCCTCGTGGAATCTACACTAGTGGGCGAGGAAGTTCGGCGGTTGGGC 3273
Db       2313 TCCATAAGATAGCACCCAGAGGTGTTTACACTAGTGGCAAGGGTTCATCTGCTGTTGGTT 2372

Qy       3274 TGACAGCGTATGTAACGAAGGATCCAGAAACTCGAGAGACGGTATTGGAGAGCGGAGCTT 3333
Db       2373 TAACTGCCTATATTACTCGTGACCAAGACACGAAGCAACTAGTCTTGGAAAGTGGTGCTT 2432

Qy       3334 TGGTTCTTAGTGATCGTGGGATATGCTGTATCGATGAGTTCGACAAAATGTCTGATAATG 3393
Db       2433 TGGTACTGTCAGATGGGGGCATTTGCTGCATTGATGAGTTTGACAAAATGTCCGATGCTA 2492

Qy       3394 CCCGAAGCATGCTTCATGAGGTAATGGAGCAACAAACGGTATCTGTAGCCAAAGGGGGTA 3453
Db       2493 CACGTTCTATTCTTCATGAAGTTATGGAGCAGCAAACCGTTACAGTTGCCAAGGCTGGCA 2552

Qy       3454 TCATTGCCTCGCTGAACGCTCGGACGTCTGTCCTTGCATGTGCAAATCCTAGTGGGTCCC 3513
Db       2553 TTATTACTACTTTGAACGCTAGGACTTCTATTTTAGCTAGTGCAAATCCAATTGGAAGCA 2612

Qy       3514 GATACAATGCGCGCCTTTCTGTGATTGATAACATCCAGCTTCCTCCAACTCTACTTTCTA 3573
Db       2613 AATACAATCCAGACTTGCCTGTCACCAAAAATATTGATCTTCCGCCTACTCTCTTATCTA 2672

Qy       3574 GATTTGATTTAATTTACTTAATGCTCGACAAACCAGACGAGCAAAACGATCGTCGTCTCG 3633
Db       2673 GATTTGATTTGGTGTATTTAATATTAGACCGCGTGGATGAAACGCTGGATAGGAAGCTCG 2732

Qy       3634 CCAGGCATCTCGTGGCTTTACACTATGAAAA------CTATGAAGTTTCAAAGCAGGACG 3687
Db       2733 CAAATCATATTGTTTCTATGTACATGGAAGATACTCCTGAACATGCTACAGATATGGAGG 2792

Qy       3688 CCTTAGATCTACAAACACTTACCGCGTATATCACCTATGCTCGTCAGCATGTACATCCTA 3747
Db       2793 TCTTTAGTGTGGAATTTTTGACAAGCTATATCACATATGCTAGAAATAATATTAATCCTG 2852

Qy       3748 CATTAAGTGATGAAGCTGCTGAAGATTTGATTAATGGCTATGTTGAGATGCGCCAAAAGG 3807
Db       2853 TCATTAGTGAAGAGGCAGCAAAGGAACTAGTGAATGCTTATGTTGGTATGCGAAAGCTTG 2912

Qy       3808 GCAACTTTCCTGGAAGCAGTAAAAAGGTGATAACAGCCACACCTCGGCAACTCGAAAGTA 3867
Db       2913 GGGAGGATGTGCGAGCATCTGAGAAAAGAATTACTGCTACAACAAGACAACTGGAATCAA 2972

Qy       3868 TGATTCGTATCAGTGAAGCCCTAGCTCGAATGAGATTTTCTGAAGTGGTAGAGAAAGTTG 3927
Db       2973 TGATTCGTCTATCTGAGGCCCATGCGAAAATGCATCTTCGCAACGTTGTTGAGGTAGGAG 3032

Qy       3928 ATGCAGCAGAAGCTGTGCGCCTTTTAGACGTCGCTTTGCAGCAATCTGCTACTGATCATG 3987
Db       3033 ATGTTCTTGAAGCTGCTCGCTTGATTAAGACCGCTATTAAGGATTATGCAACTGATCCTG 3092

Qy       3988 CAACAGGTACGATAGACATGGATCTTAT 4015
Db       3093 CAACAGGAAAGATTTCTCTTGACCTAAT 3120

RESULT 6
AC116551/c
LOCUS       AC116551              153751 bp    DNA     linear   INV 21-FEB-2004
DEFINITION  Dictyostelium discoideum chromosome 2 map complement(581327-427576)
            strain AX4, complete sequence.
ACCESSION   AC116551 AC116101 AC116549 AC116965
VERSION     AC116551.2  GI:28829639
KEYWORDS    HTG.
SOURCE      Dictyostelium discoideum
  ORGANISM  Dictyostelium discoideum
            Eukaryota; Mycetozoa; Dictyosteliida; Dictyostelium.
REFERENCE   1  (bases 1 to 153751)
  AUTHORS   Gloeckner,G., Eichinger,L., Szafranski,K., Pachebat,J., Dear,P.,
            Lehmann,R., Baumgart,C., Parra,G., April,J.F., Guigo,R., Kumpf,K.,
            Tunggal,B., Cox,E., Quail,M.A., Platzer,M., Rosenthal,A. and
            Noegel,A.A.
  TITLE     Sequence and analysis of chromosome 2 of Dictyostelium discoideum
  JOURNAL   Nature 418 (6893), 79-85 (2002)
   PUBMED   12097910
  REMARK    The Dictyostelium Genome Sequencing Consortium
REFERENCE   2  (bases 1 to 153751)
  AUTHORS   Baumgart,C.
  TITLE     Direct Submission
  JOURNAL   Submitted (29-MAR-2002) Genome Analysis, Institute of Molecular
            Biotechnology, Beutenberstr. 11, Jena 07745, Germany
REFERENCE   3  (bases 1 to 153751)
  AUTHORS   Baumgart,C.
  TITLE     Direct Submission
  JOURNAL   Submitted (04-MAR-2003) Genome Analysis, Institute of Molecular
            Biotechnology, Beutenberstr. 11, Jena 07745, Germany
REFERENCE   4  (bases 1 to 153751)
  AUTHORS   Baumgart,C.
  TITLE     Direct Submission
  JOURNAL   Submitted (12-MAR-2003) Genome Analysis, Institute of Molecular
            Biotechnology, Beutenberstr. 11, Jena 07745, Germany
REFERENCE   5  (bases 1 to 153751)
  AUTHORS   Baumgart,C.
  TITLE     Direct Submission
  JOURNAL   Submitted (21-FEB-2004) Genome Analysis, Institute of Molecular
            Biotechnology, Beutenberstr. 11, Jena 07745, Germany
COMMENT     On or before Feb 21, 2004 this sequence version replaced
            gi:28850309, gi:19807752, gi:19920064, gi:19807784.
            CDS predictions from GeneID do not necessarily reflect true genes.
            Further Information is available from IMB Jena, Department of
            Genome Analysis
            (http://genome.imb-jena.de/dictyostelium/)
            and the Univerity Colonge, Institute for Biochemistry I
            (http://www.uni-koeln.de/dictyostelium/project.shtml
            Funding
            Agency : Deutsche Forschungsgemeinschaft (DFG).
FEATURES             Location/Qualifiers
     source          1..153751
                     /organism="Dictyostelium discoideum"
                     /mol_type="genomic DNA"
                     /strain="AX4"
                     /db_xref="taxon:44689"
                     /chromosome="2"
                     /map="complement(581327-427576)"
     CDS             complement(join(<351..2554,2687..2807))
                     /note="GeneID exon scores (in order of location ranges):
                     70.76, 0.70 - GSCJ_ID dd_02254"
                     /codon_start=1
                     /product="hypothetical protein"
                     /protein_id="AAO52218.1"
                     /db_xref="GI:28829702"
                     /translation="MNESISDHNEKLFWKVFRNKYLNIEIFKQLESTEIEYDSFTIAS
                     LIKLSVNNSNNNNDNNNKLIEIFKLLMEKEKYLMEGLDLVTISISSRNFEILKILLNK
                     PYSFLVKETDFNDALQKYGDSEIEKSILQLLGNQININTFLIENNFQIKNSIPFFKNT
                     IQFILNNQNNDDNNDDNKNNNNKLIKLLELKIYSIKKDEILNLELIEMQNKLLDLNYF
                     SKENFQNSNIYWFKELYEFKGVKTIFIHLILLVFKIYYKRKSLNIPNEIEIEINEIMN
                     NSLLIKNEEKEKKLLKLLINETQSNLLFKLYYYRYYEIIEGCKLSKIVYRNLKTFNIF
                     KLIESNNNSNNNKEEIIRIIKNFINVYEFKNNENDKQLINKFRNVILNNLNSKEYYHK
                     LIETIGNSQIKTFGLPIEWNINLLDFLLDDEINNKIIIIKDKQVIDWLSNMPRLNNRI
                     FTFDIKTIYYYKRKKNLIQFQSLELIEYAIKKLETNNNNNSNNDNLPPIKLETINWSM
                     VYNNPLKSIIDEFKFQNESIINEDELNDKVTFNLFKFLKFEKNEFTILVQLIKYFIEK
                     VLNKYGQDKSYEILKNFVKIDNLCKFQKETLELLVLSTINLDLIMVMIILNCGIDFTN
                     KQDEILINFWNVHILTNTSIFYIPTINGNINFFKIITHIKPQYIINNIKLWCEQTSLL
                     ECKHLAINHLFKKLIQTKGTTLNQILYIKDFINTINNNNNNNNNNNNNNNIIKLDHS
                     KFHSIHYLNLKSPKLILYLIESNQIRDWNYLLIPQ"
     CDS             3515..5095
                     /note="GeneID exon scores (in order of location ranges):
                     144.31 - GSCJ_ID dd_02253"
                     /codon_start=1
                     /product="similar to Arabidopsis thaliana (Mouse-ear
                     cress). Hypothetical 59.5 kDa protein"
                     /protein_id="AAO52156.1"
                     /db_xref="GI:28829640"
```

```
                     /translation="MNNQVKPTLSNKAEIINKWLIDNGVRINNKLIKIVYLGKENNFE
                     QTENTTATTTTSERINDSIVSGLGVISLKELKVDDIVAKIPKSIILSIHTSSISNILE
                     KYKIENNIGTSIALIHEASLGEKSKWYGYISSLPRKVDVPILWDSESRKLLKGTAIED
                     VLNDDDILINQVYADVIESILSKNHPEIFGDKELYSIENFKIANSIISSRAFCVDSYH
                     GDSLVPLADIFNHQTAREHVHIESNGDVCNKCGSIKTCKHRKVVTQHHTVNSTKGKRT
                     HKVAGIPSSKKHIHKGNCCSTTTTKTNEEDKDTIIEEDDEHLYIKVVKGVEANKEVYN
                     TYGDHDNAILLSKYGFLEMDNPCDRLSIDKQLVDKELELIGAENQITRMDLFNRISFY
                     AEMFDIDSRNNHAIENDGRLDDALVCSVGIALAPQSIFQSWKSMSEHKREKYFEQLEA
                     EDIVKQSQLVRNTLLSILNKLFSNYNNSNETTATTTIEQDQEKLSTLKNQREIIAQSL
                     KICEKTLINKSINYYKNL"
     CDS             complement(5235. .7895)
                     /note="GeneID exon scores (in order of location ranges):
                     270.65 - GSCJ_ID dd_03185"
                     /codon_start=1
                     /product="similar to essential for initiation of DNA
                     replication; homolog of S. pombe CDC21; Cdc54p
                     [Saccharomyces cerevisiae]"
                     /protein_id="AAO52157.1"
                     /db_xref="GI:28829641"
                     /translation="MDPKETKTSKISFLNTTKKNNNNNNNNNNNNNNNNNNNNNNNNN
                     NNNNNNNNNDDNNNNGENNIDIEPDDIDLLNKNDIDNIVDGIDNNNKSGSNDMKGLVF
                     WGTTVKVEDVKERFRKFIFNFPTPKERKRFNINDINNNNNNNNNNNSEEMNDDDDYYY
                     DEEQDKLIKKPEKVQYLYRDLLKVLNETKKCHLNINMKYLVQFDMELYLQWVSFPNEM
                     IPLLDEEINLIYRELFPITPRNGDDEDDDDDDDEDEEDYRIELHPFNLLRKTPMRDLN
                     PSDIDKIISISGLIIRSSSIIPEIKQAFFMCAVCEATFHANVEKGKIQEPSECSNCKS
                     KQSLSIIHNRCLFGDKQYIKLQETPDAIPEGETPHTVALFAYGDLIDIAKPGDRVELT
                     GVFKASPMRAGSNRSLRSIYKTYIDILHIKRTDKGKYDDDDDDHDDNTGGGGTGTGKE
                     TNENLDFEDLDEFDLSEEKEREIIELSKKPDIYDIVTKSIAPNIWELEDIKKGILCQL
                     FGGSKKSYQDYGGKFRGDINILLCGDPGTSKSQLLSYVHKIAPRGIYTSGKGSSAVGL
                     TAYITKDPDTKETVLESGALVLSDKGVCCIDEFDKMNDQTRSILHEVMEQQTVSIAKA
                     GIICTLNARTSILASANPSGSRYMPKLSVVENIQLPPTLLSRFDLIYLVLDKANERSD
                     RQLARHLVSMYWDETPVSHFTIPKETLTNYIQYARKHINPKLTDDSAKCLVQGYLEMR
                     SMGSSKKTISATPRQLESLIRIAEAHARIRFSEFVEPLDVEEAIRLIKVALQQAAIDP
                     ENGTIDMDLITTGRSASSREAITRLKSHIKQKIGKKHLTLDQLLKLLTIQNQVQTIQQ
                     IEEIKEALRQLQDEEIIQSSGGIDPTYGVV"
     CDS             join(8672. .9027,9184. .9463)
                     /note="GeneID exon scores (in order of location ranges):
                     37.11, 23.19 - GSCJ_ID dd_02251"
                     /codon_start=1
                     /product="similar to cyctochrome c1 heme lyase
                     [Caenorhabditis elegans]"
                     /protein_id="AAO52158.1"
                     /db_xref="GI:28829642"
                     /translation="MGDDNTNNTSITKSSPTDAMGGGCPIVHDINPANHMYKPNQNPH
                     PDQTKPLSTERITSTIPRTEKDNWQYPSPQMFFNAMKKKQYEPKEEDMSVVISIHNTV
                     NEKCWEDVLQWENDYKDVCPNPKLKKFKGKATDFSPKAKFLNTFLGYKLPFDRHDWIV
                     DRNGKEVRYVIDFYEGRINKDSGKSIGIYIDVRPAIDDFSSLKDRVLHFFK"
     CDS             complement(join(9650. .13297,13390. .13536))
                     /note="GeneID exon scores (in order of location ranges):
                     468.70, 11.48 - GSCJ_ID dd_02250"
                     /codon_start=1
                     /product="similar to C25A11.4b.p [Caenorhabditis elegans]"
                     /protein_id="AAO52159.1"
                     /db_xref="GI:28829643"
                     /translation="MNPFEMIQNQNIVIKNAIEVGCIIVNIGAQDLMNATPYLVLGVI
                     WQIIKQGLLSKVNQNANEILEILFEEDSVKEVDNTQNQSDEHSAEEILLRWVNYHLEK
                     EGCERRISNFSEDIQDSVVYSHLFHQLVPIEFENLVSTIHSESNLFSRAELITKACEK
                     IGVKCFLTPSDIALGHPKLNLALVASLFNSEAAIANLRKEREEKHRTDQERLERERLE
                     RERIERERLERERLEQERLERERLERERLEKQRQEMERIEQERIELERLERERLEQER
                     LEQERIERERLERERLEQERLEQERLERERLEQERLECERLERERLEQERLEQERLER
                     ERIELEELERIERERLEYEERERIERERIEREREELEEQERLEYERLEQERIEREEME
                     RIERERIEREEFERLEEERIERERERALRLEAERLERELRERERITRQSELERIEYNK
                     EYKICIRIQKRMIDDTSLTDNMTKLIGDITELKRNIIGELRNTTHHQAELKKLTKSID
                     RVLENKNEGKKLKKQKKVIIKPVKGVPNGLDPNKMKNYQSLFYYLQTRPEVMGKLLFL
                     LDRNQMGKTFDNLLLSMFPYEISPREEDLFLKALTVGIEYQIKYSETLQDFLQSESVP
                     MKLLYQYFKKKGFKYLKEIITPLILKVLDQTDLNLNFDPYWIQKEIQIAKELKDGKKS
                     EKKTLDQTFEQLMANNDVVSIFNSRCLQLVELGNAFLERITKSIEQLPVQLRYLCRII
                     AEKSKYAFPTNNTNNHHNGGGISSDESGSTTANITLNNNRSKQLPENREKYISYFLII
                     YRFIIPIISSPDLLNELVLLQHPYITKQSRFNLSQLSRLIQSVFSEKPWMNYFPTVYL
                     NKWIIQNSSSADKFLCQVAQVPDLDELIHYSEYNEIMLENNIKFCSVINISEIIWLHE
                     TILSNFYDLEDIGTLKEVSEKSIQNLAERAINSKLDKGKYVTAARQSTPAIMTSQKSQ
                     NWPLKKYLDEIDFPIQLPEQGDRSIQLSFINRLDQLNEQSIDLWKECKLWMIDILSKS
                     QPPKKNANTIEIIREAIDYFKTGSNSSENLSIINKSNELISRLRTFDRDQCTDGLDGG
                     GPDNFNLFLLDVEKELATILRKNKVMRSELTRLHSLLKSIRIDAAPLEHLTKFQNEQL
                     EKAIVKKQLKTTQLTPLPKDKSTIKPHKFILLDLVKKGMILRCELPESSFSKTKIYIK
                     SLGPGLFEIEAKFGPITKSLQLELDDLLEKSKNHIDEFPLEGITLDVTLTIHLLMKLF
                     SL"
     CDS             complement(join(13685. .13812,14033. .14207))
                     /note="GeneID exon scores (in order of location ranges):
                     6.21, 14.57 - GSCJ_ID dd_02249"
                     /codon_start=1
                     /product="hypothetical protein"
                     /protein_id="AAO52160.1"
                     /db_xref="GI:28829644"
                     /translation="MTYTDSRYNTLNRGSKKMDPPFITQEVLASAKISKKDNILINAN
                     AEDGSVHTFSEEEKVAYSKFITERLIDQEDVLSNYVPIDPSTDQLFVNCTDGVLLK"
     CDS             16650. .17525
                     /note="GeneID exon scores (in order of location ranges):
                     78.29 - GSCJ_ID dd_00829"
                     /codon_start=1
                     /product="hypothetical protein"
                     /protein_id="AAO52161.1"
                     /db_xref="GI:28829645"
                     /translation="METKDKKKDMELLGDRMKSYEDDMKIQIEKNKPFIIRLDGHSFS
                     KFTKNFNKPHDIRIHNAMIETSTVLLKTFMPTCIYTFSDEITMCFPSIDESTLEEGKE
                     IPNLAYSGKVQKLISLSSGLASTVFFKSITNAQYDKDTELNLIKLLETCTPHFDARIF
                     TLPSNQEIVNNLIWRSLVDCKRNSVANLGFAHFTPKQMLGLNNTEVKKKLLEEKSIIY
                     ENEPAWYRFGTYLKKEYYTLTTVSPNEPEKQITAIRSKVNCFSFDITKLSNSLNFISF
                     KTLPEGFNLELIESK"
     CDS             complement(join(17735. .18959,19168. .19640,19750. .20190,
                     20286. .20531))
                     /note="GeneID exon scores (in order of location ranges):
                     130.29, 48.89, 51.03, 20.75 - GSCJ_ID dd_00831"
                     /codon_start=1
                     /product="similar to Dictyostelium discoideum (Slime
                     mold). ALG-2 interacting protein X"
                     /protein_id="AAO52162.1"
                     /db_xref="GI:28829646"
                     /translation="MLSIERKRTEKVDFSKPLTKYIKEQFSKAESDQHETQIATLNGL
                     REDVRNLQERTETSKEMVWKYYSILSSLELRFPISENNVRISFPWTDSYRQRKSTLYS
                     IYFERASVLFNYGSIVSQIASSTNRSNIEGVKKACNQFQLAAGVFNKLREYASLHPEC
                     STSADFSSESLQALVTIMLAQAQECIYEKASMDNLSDSILSKLAAQVAEYYDTFNQLL
                     NSNSLKSIVDRNWNITATVKSYLYKAISLYCHAKGLEQVSQFGEQVSRLMIAVDNINQ
                     SKVNLAKTAPIELKEIVERYVISITRYCESAKKDNDTIYHDTIPPAHKLTPIEKKPLA
                     KALPLPEINFVDPFNSLVPFSVKEDSAYYNDQKETLLRKELDNIEFHNQSAKASLLSM

  Query Match             11.3%;  Score 492.4;  DB 2;  Length 153751;
  Best Local Similarity   59.3%;  Pred. No. 6.6e-135;
  Matches 1011;  Conservative    0;  Mismatches  636;  Indels   57;  Gaps    8;

Qy       2370 ATTGAGGCCAGACCTTTCAATCTCAAAGCATCGGTGCACATGCGTGAACTCAACCCTTCA 2429
              |||||      || |||||| |   |  |      |  ||| | ||  | || || |||
Db       7121 ATTGAATTACATCCATTCAATTTATTAAGAAAAACACCAATGAGAGATTTAAATCCATCA 7062

Qy       2430 GATATAGACAAATTGGTTTCTGTTAAAGGAATGGTTATCCGGTGCAGTTCTATCATACCT 2489
              ||||| || ||   |  | ||  |||  ||  |  ||||  |  | ||||| || |||||
Db       7061 GATATTGATAAGATTATATCAATTAGTGGTTTAATTATTAGAAGTAGTTCAATTATACCA 7002

Qy       2490 GAAATTAAGGGGGCCTTCTTCAAATGTTTAGTGTGTGGTCACTCGCCTCCGCTAGTTACA 2549
              || |||||    || || || |  |||  ||| ||||    |  |  ||   |  |  |
Db       7001 GAGATTAAACAAGCGTTTTTTATGTGTGCAGTTTGTG---AAGCAACTTTTCATGCAAAT 6945

Qy       2550 GTTGTTAAAGGGCGGGTTGAGGAGCCAACAAGGTGTGAAAAGCCAGAATGTGCAGCACGG 2609
              ||||  || ||    || | ||||||| ||     |||    || | |||  | ||   |
Db       6944 GTTGAGAAGGGTAAAATTCAAGAGCCATCA------GAATGTTCAAATTGTAAATCAAAG 6891

Qy       2610 AATGCTATGTCTCTTATTCACAATCGATGCACTTTTGCAAATAAGCAGATAGTGCGTCTT 2669
               |  |  | || | ||||| |||||| ||    ||||   |||| |||     |    |||
Db       6890 CAATCATTATCAATCATTCATAATCGTTGTTTGTTTGGTGATAAACAGTATATTAAACTT 6831

Qy       2670 CAAGAAACTCCAGATGCCATTCCTGAAGGAGAGACTCCACACACAGTCAGCATGTGTTTA 2729
              |||||||| |||||||| ||||| || || || || ||||| || |||      | | |
Db       6830 CAAGAAACACCAGATGCAATTCCAGAGGGTGAAACACCACATACCGTCGCATTATTTGCC 6771

Qy       2730 TACAACACTATGGTTGATGCTGTGAAGCCTGGAGATCGTATTGAGGTAACAGGAGTTTTC 2789
              ||        | |  ||||| ||  || || || ||||||| ||| |||| || || ||
```

```
Db     6770 TATGGTGATTTAATTGATATTGCAAAACCAGGTGATCGTGTCGAGTTAACTGGTGTATTT 6711

Qy     2790 AAGGCCATGGCAGTTCGAGTTGGTCCGAATCAACGAACATTACGAGCATTGTATAAGACC 2849
Db     6710 AAAGCAAGTCCAATGAGAGCTGGTTCAAAT---AGATCACTTAGAAGTATTTATAAAACT 6654

Qy     2850 TACATCGATTGCGTGCACGTCAAGAAGTCTGACAGGGGTCGACTGCAAACTGAAGATCCT 2909
Db     6653 TATATTGATATTTACATATAAAAAGAACAGATAAAGGTAAATATGATGATGATGATGAT 6594

Qy     2910 ATGGAGATGGATAA------GGAGAATGATATGTATGCTGGGTATCATGAAAGTGATACT 2963
Db     6593 GATCATGATGATAATACTGGTGGTGGTGGCACAGGTACAGGTAAAGAAACAAATGAAAAT 6534

Qy     2964 TCAGAAGCTGCTAA--------------------TGAAGCAAAGATTCAAAAACTTAAA 3002
Db     6533 TTAGATTTTGAAGATTTAGATGAATTTGATTTATCTGAAGAAAAAGAAAGAGAGATTATA 6474

Qy     3003 GAGCTGTCCAAGCTCCCGGGCATTTATGATAGACTTTCAAGGTCGCTGGCTCCAAGCATT 3062
Db     6473 GAATTATCAAAAAAACCAGACATTTATGATATTGTAACTAAATCAATAGCACCAAATATT 6414

Qy     3063 TGGGAGCTTGAAGATATTAAAAAGGGTCTTCTTTGCCAGCTCTTTGGTGGGAAGGCTAAG 3122
Db     6413 TGGGAATTAGAAGATATTAAAAAAGGTATACTTTGTCAATTATTTGGTGGTTCAAAAAAG 6354

Qy     3123 AAAATTCCA---TCTGGAGCATCTTTCCGAGGTGACATCAATGTTTTACTTGTTGGGGAC 3179
Db     6353 AGTTATCAAGATTATGGTGGAAAGTTTAGAGGTGATATAAACATTTTATTGTGTGGTGAT 6294

Qy     3180 CCTGGTACCAGTAAATCTCAGCTGCTTCAGTATGTGCACAAGATAGCTCCTCGTGGAATC 3239
Db     6293 CCTGGTACTTCGAAATCTCAATTACTCTCGTATGTTCATAAAATTGCACCCCGTGGTATC 6234

Qy     3240 TACACTAGTGGGCGAGGAAGTTCGGCGGTTGGGCTGACAGCGTATGTAACGAAGGATCCA 3299
Db     6233 TATACAAGTGGTAAGGGTAGTAGTGCAGTTGGTTTGACGGCATATATCACCAAAGATCCC 6174

Qy     3300 GAAACTCGAGAGACGGTATTGGAGAGCGGAGCTTTGGTTCTTAGTGATCGTGGGATATGC 3359
Db     6173 GATACTAAAGAGACCGTTTTAGAGAGTGGTGCATTGGTACTCTCTGATAAGGGTGTTTGT 6114

Qy     3360 TGTATCGATGAGTTCGACAAAATGTCTGATAATGCCCGAAGCATGCTTCATGAGGTAATG 3419
Db     6113 TGTATTGATGAATTTGATAAAATGAATGATCAAACTCGTTCAATTCTTCATGAAGTTATG 6054

Qy     3420 GAGCAACAAACGGTATCTGTAGCCAAGGGGGTATCATTGCCTCGCTGAACGCTCGGACG 3479
Db     6053 GAACAGCAAACAGTATCAATTGCAAAGGCGGGCATTATCTGCACGTTGAATGCCAGAACC 5994

Qy     3480 TCTGTCCTTGCATGTGCAAATCCTAGTGGGTCCCGATACAATGCGCGCCTTTCTGTGATT 3539
Db     5993 TCGATTCTTGCATCTGCAAATCCAAGTGGTTCACGTTATATGCCCAAATTATCAGTGGTG 5934

Qy     3540 GATAACATCCAGCTTCCTCCAACTCTACTTTCTAGATTTGATTTAATTTACTTAATGCTC 3599
Db     5933 GAGAATATCCAATTGCCTCCAACCCTATTATCGAGATTCGATTTAATTTATTTAGTATTG 5874

Qy     3600 GACAAACCAGACGAGCAAAACGATCGTCGTCTCGCCAGGCATCTCGTGGCTTTACACTAT 3659
Db     5873 GATAAAGCCAATGAAAGAAGTGATCGTCAATTGGCTCGTCATTTAGTATCGATGTATTGG 5814

Qy     3660 GAAAACTATGAAGTTTCAAAGCAGGACGCCTTAGATCTACAAACACTTACCGCGTATATC 3719
Db     5813 GATGAAACACCGGTTTC---TCATTTCACAATCCCCAAGGAAACCCTCACAAACTATATT 5757

Qy     3720 ACCTATGCTCGTCAGCATGTACATCCTACATTAAGTGATGAAGCTGCTGAAGATTTGATT 3779
Db     5756 CAATACGCTAGAAAACATATTAATCCAAAGTTAACTGATGATTCTGCAAAATGTTTAGTT 5697

Qy     3780 AATGGCTATGTTGAGATGCGCCAAAAGGGCAACTTTCCTGGAAGCAGTAAAAAGGTGATA 3839
Db     5696 CAAGGTTATTTAGAGATGAGAAGTATGGGT------------AGTAGTAAAAAAACAATC 5649

Qy     3840 ACAGCCACACCTCGGCAACTCGAAAGTATGATTCGTATCAGTGAAGCCCTAGCTCGAATG 3899
Db     5648 TCTGCCACACCAAGACAATTGGAATCACTTATTCGTATCGCTGAAGCTCATGCTAGAATT 5589

Qy     3900 AGATTTTCTGAAGTGGTAGAGAAAGTTGATGCAGCAGAAGCTGTGCGCCTTTTAGACGTC 3959
Db     5588 CGTTTCTCTGAATTTGTTGAACCATTAGATGTTGAAGAAGCAATTCGTTTAATTAAAGTC 5529

Qy     3960 GCTTTGCAGCAATCTGCTACTGATCATGCAACAGGTACGATAGACATGGATCTTATCACG 4019
Db     5528 GCACTTCAACAAGCTGCAATCGATCCTGAAAATGGTACAATCGATATGGATCTAATTACA 5469

Qy     4020 ACTGGAGTGTCGGCCAGCGAGCGT 4043
Db     5468 ACTGGTCGTTCTGCAAGTTCTCGT 5445
```

```
RESULT 7
AX489180
LOCUS       AX489180                2733 bp    DNA     linear   PAT 16-AUG-2002
DEFINITION  Sequence 6480 from Patent WO02053728.
ACCESSION   AX489180
VERSION     AX489180.1  GI:22323192
KEYWORDS    .
SOURCE      Candida albicans
  ORGANISM  Candida albicans
            Eukaryota; Fungi; Ascomycota; Saccharomycotina; Saccharomycetes;
            Saccharomycetales; mitosporic Saccharomycetales; Candida.
REFERENCE   1
  AUTHORS   Roemer,T., Jiang,B., Boone,C., Bussey,H. and Ohlsen,K.L.
  TITLE     Gene disruption methodologies for drug target discovery
  JOURNAL   Patent: WO 02053728-A 6480 11-JUL-2002;
            Elitra Pharmaceuticals, Inc. (US)
FEATURES             Location/Qualifiers
     source          1..2733
                     /organism="Candida albicans"
                     /mol_type="unassigned DNA"
                     /db_xref="taxon:5476"
ORIGIN

  Query Match             11.0%;  Score 480.2;  DB 6;  Length 2733;
  Best Local Similarity   56.6%;  Pred. No. 1.7e-131;
  Matches  959;  Conservative    0;  Mismatches  718;  Indels   18;  Gaps    3;

Qy     2409 ATGCGTGAACTCAACCCTTCAGATATAGACAAATTGGTTTCTGTTAAAGGAATGGTTATC 2468
Db      883 ATTCGTGAATTGAATCCTAATGATATTGATAAGTTAGTTAGTGTTAAAGGGTTAACATTA 942

Qy     2469 CGGTGCAGTTCTATCATACCTGAAATTAAGGGGGCCTTCTTCAAATGTTTAGTGTGTGGT 2528
Db      943 CGTTCTACTTCTATCATTCCTGATATGAAAGTGGCATTTTTCAGATGTAACGCCTGTGGT 1002

Qy     2529 CACTCGCCTCCGCTAGTTACAGTTGTTAAAGGGCGGGTTGAGGAGCCAACAAGGTGTGAA 2588
Db     1003 CATACTGTTGGG---GTGGAAATTGATCGTGGTGTAATTTCCGAACCAACAAAATGTCCT 1059

Qy     2589 AAGCCAGAATGTGCAGCACGGAATGCTATGTCTCTTATTCACAATCGATGCACTTTTGCA 2648
Db     1060 AGAGAAGTATGTGGTCAAACCAATTCAATGGTACTTATTCATAATCGTTCATCATTTTCT 1119

Qy     2649 AATAAGCAGATAGTGCGTCTTCAAGAAACTCCAGATGCCATTCCTGAAGGAGAGACTCCA 2708
Db     1120 GATAAACAAGTGATTAAATTACAAGAAACTCCTGATTTGGTACCTGATGGTCAAACCCCT 1179

Qy     2709 CACACAGTCAGCATGTGTTTATACAACACTATGGTTGATGCTGTGAAGCCTGGAGATCGT 2768
Db     1180 CATTCTATCAATTTATGTGTTTACGATGAATTGGTTGATTCTTGTCGTGCTGGTGATCGT 1239

Qy     2769 ATTGAGGTAACAGGAGTTTTCAAGGCCATGGCAGTTCGAGTTGGTCCGAATCAACGAACA 2828
Db     1240 GTTGAAGTATGTGGTATTTTCAGATCAACTCCAGTTCGAGCAAATCCTAGACAAAGAGCT 1299
```

```
Qy       2829 TTACGAGCATTGTATAAGACCTACATCGATTGCGTGCACGTCAAGAAGTCTGACAGGGGT 2888
Db       1300 TTAAAAAACTTATATAAAACATATTTGGATATTGTTCATGTGAAAAAAATCGATAAAAGG 1359

Qy       2889 CGACTGCAAACTGAAGATCCTATGGAGATGGATAAGGAGAATGATATGTATGCTGGGTAT 2948
Db       1360 CGATTAGGTGGTGATGTTACCA---------CTTTAGAACACGAATTGGCTGAAAAGGAT 1410

Qy       2949 CATGAAAGTGATACTTCAGAAGCTGCTAATGAAGCAAAGATTCAAAAACTTAAAGAGCTG 3008
Db       1411 CAAGAAGTTGAACAAGTTCGAAAAATCACTGCTGAAGAAGAAGCTAAAATCAAAGAAATT 1470

Qy       3009 TCCAAGCTCCCGGGCATTTATGATAGACTTTCAAGGTCGCTGGCTCCAAGCATTTGGGAG 3068
Db       1471 TCTGAACGTGATGATTTGTATGAAATATTGGCTAGATCATTAGCTCCATCTATTTATGAG 1530

Qy       3069 CTTGAAGATATTAAAAAGGGTCTTCTTTGCCAGCTCTTTGGTGGGAAGGCTAAGAAAATT 3128
Db       1531 ATGGATGATGTTAAAAAGGGGATCTTGTTACAATTATTTGGTGGTACTAATAAAACATTT 1590

Qy       3129 CCATCTGGAGCATCTTTCCGAGGTGACATCAATGTTTTACTTGTTGGGGACCCTGGTACC 3188
Db       1591 ACAAAAGGTGGTCGTTATAGAGGTGATATAAATATTTTATTATGTGGTGATCCATCTACT 1650

Qy       3189 AGTAAATCTCAGCTGCTTCAGTATGTGCACAAGATAGCTCCTCGTGGAATCTACACTAGT 3248
Db       1651 TCTAAATCACAAATCTTACAATATGTTCATAAGATTGCTCCAAGAGGGGTTTATACTTCT 1710

Qy       3249 GGGCGAGGAAGTTCGGCGGTTGGGCTGACAGCGTATGTAACGAAGGATCCAGAAACTCGA 3308
Db       1711 GGTAAAGGTTCTTCGGCTGTTGGTTTGACAGCTTATATCACCAGAGATATTGATACAAAA 1770

Qy       3309 GAGACGGTATTGGAGAGCGGAGCTTTGGTTCTTAGTGATCGTGGGATATGCTGTATCGAT 3368
Db       1771 CAATTAGTATTGGAAAGTGGGGCCCTTGTGTTATCTGATGGAGGTGTTTGTTGTATTGAT 1830

Qy       3369 GAGTTCGACAAAATGTCTGATAATGCCCGAAGCATGCTTCATGAGGTAATGGAGCAACAA 3428
Db       1831 GAGTTTGACAAGATGAGTGATGCCACCAGATCAGTTTTACATGAAGTCATGGAACAACAA 1890

Qy       3429 ACGGTATCTGTAGCCAAAGGGGGTATCATTGCCTCGCTGAACGCTCGGACGTCTGTCCTT 3488
Db       1891 ACTATTTCTATTGCTAAAGCCGGTATTATTACCACATTAAATGCTAGGACATCTATTTTA 1950

Qy       3489 GCATGTGCAAATCCTAGTGGGTCCCGATACAATGCGCGCCTTTCTGTGATTGATAACATC 3548
Db       1951 GCTTCTGCAAATCCAATTAATTCTCGTTATGATCCAAATTTGCCTGTGACGGGTAACATT 2010

Qy       3549 CAGCTTCCTCCAACTCTACTTTCTAGATTTGATTTAATTTACTTAATGCTCGACAAACCA 3608
Db       2011 GATTTGCCACCACCATTGTTATCACGTTTTGATTTGGTTTATTTGATTTTGGATAAAGTT 2070

Qy       3609 GACGAGCAAAACGATCGTCGTCTCGCCAGGCATCTCGTGGCTTTACACTATGAAAACTAT 3668
Db       2071 GATGAATCAATCGATAGACAATTGGCTAGACATTTGACTGATATGTATTTGGAAGATGCT 2130

Qy       3669 GAAGTTTCAAAGCAGGACGCCTTAGATCTAC------AAACACTTACCGCGTATATCACC 3722
Db       2131 CCTGAAACTGTTAATGCCAATTCGGTATTACCCGTGGAATTATTAACACTTTATATTCAA 2190

Qy       3723 TATGCTCGTCAGCATGTACATCCTACATTAAGTGATGAAGCTGCTGAAGATTTGATTAAT 3782
Db       2191 TATGCTAAAGAGAATTTCAATCCCGTGATGACAGAAGAAGGTAAAAATGAATTAGTTAGA 2250

Qy       3783 GGCTATGTTGAGATGCGCCAAAAGGGCAACTTTCCTGGAAGCAGTAAAAAGGTGATAACA 3842
Db       2251 TCATACGTTGAAATGAGAAAATTGGGGGAGGATGCTAGATCATCAGAAAAGAGAATCACC 2310

Qy       3843 GCCACACCTCGGCAACTCGAAAGTATGATTCGTATCAGTGAAGCCCTAGCTCGAATGAGA 3902
Db       2311 GCTACTACAAGACAATTGGAATCGATGATTAGATTGTCTGAAGCTCATGCTAAAATGAGA 2370

Qy       3903 TTTTCTGAAGTGGTAGAGAAAGTTGATGCAGCAGAAGCTGTGCGCCTTTTAGACGTCGCT 3962
Db       2371 TTATCTGAAAGAGTTGAATTGATCGATGTTAAAGAAGCAGTAAGATTAATTAAATCGGCC 2430

Qy       3963 TTGCAGCAATCTGCTACTGATCATGCAACAGGTACGATAGACATGGATCTTATCACGACT 4022
Db       2431 ATTAAAGATTATGCTACTGATCCAGTTACTGGTAGAATCGATATGGATATGATTCAAACT 2490

Qy       4023 GGAGTGTCGGCCAGCGAGCGTATTCGTCGGGCCAACTTGCTAGCTGCTCTGCGAGAGCTT 4082
Db       2491 GGTACTACGGCACAACAAAGAAGAGTACAAGAAGATTTGGTCAGTGAAATTATGAAAATC 2550

Qy       4083 ATAGCAGATAAAATT 4097
Db       2551 ATCGAAGAAAATAAT 2565

RESULT 8
CR954214_1/c
WPCOMMENT
  Sequence split into 6 fragments  LOCUS CR954214 Accession CR954214
     Fragment Name      Begin      End
     CR954214_0             1      110000
     CR954214_1        100001      210000
     CR954214_2        200001      310000
     CR954214_3        300001      410000
     CR954214_4        400001      510000
     CR954214_5        500001      516206
Continuation (2 of 6) of CR954214 from base 100001 (CR954214 Ostreococcus tauri strain

  Query Match            10.8%;  Score 471.4;  DB 14;  Length 110000;
  Best Local Similarity  57.0%;  Pred. No. 1.2e-128;
  Matches  968;  Conservative    0;  Mismatches  681;  Indels   48;  Gaps    4;

Qy       2380 GACCTTTCAATCTCAAAGCATCGGTGCACATGCGTGAACTCAACCCTTCAGATATAGACA 2439
Db      72194 GCCCGTTCAACATGATGGAGGCGAAACCCATGCGCGACTTGAATCCATCGGACATCGACA 72135

Qy       2440 AATTGGTTTCTGTTAAAGGAATGGTTATCCGGTGCAGTTCTATCATACCTGAAATTAAGG 2499
Db      72134 AGATGGTGTGCGTGCGTGGTATGGTGACGCGATGCACAACGATCATTCCTGACCTAAAAC 72075

Qy       2500 GGGCCTTCTTCAAATGTTTAGTGTGTGGTCACTCGCCTCCGCTAGTTACAGTTGTTAAAG 2559
Db      72074 TCGCGTATTTCAAGTGTCTGATGTGCGGCTTCGCCCCAGAACATATTCAAGTCGACCGCG 72015

Qy       2560 GGCGGGTTGAGGAGCCAACAAGGTGTGAAAAGCCAGAATGTGCAGCACGGAATGCTATGT 2619
Db      72014 GTCGCGTCAACGAACCTCC---GCTCAAGTGCACAGAGTGCGGCAAGCCAGGAACGATGA 71958

Qy       2620 CTCTTATTCACAATCGATGCACTTTTGCAAATAAGCAGATAGTGCGTCTTCAAGAAACTC 2679
Db      71957 CCTTGATTCACAATCAATGCGTCTTCGCAAACAAGCAAACGGTGAAGATGCAGGAGACAC 71898

Qy       2680 CAGATGCCATTCCTGAAGGAGAGACTCCACACACAGTCAGCATGTGTTTATACAACACTA 2739
Db      71897 CGGATGCGATTCCCGAGGGTGAGACTCCGCACACTGTATCCATGTGCGTCTTCGATGATC 71838

Qy       2740 TGGTTGATGCTGTGAAGCCTGGAGATCGTATTGAGGTAACAGGAGTTTTCAAGGCCATGG 2799
Db      71837 TCGTTGATCAGGCGAAGCCAGGTGACCGAGTTGAAGTTACTGGAGTGTATCGTGCGGTTC 71778

Qy       2800 CAGTTCGAGTTGGTCCGAATCAACGAACATTACGAGCATTGTATAAGACCTACATCGATT 2859
Db      71777 CAATTCGTCTGAGCTCTACGAAGCGCACGCTCAAGAGTGTGTACAAGACGTACCTAGACG 71718

Qy       2860 GCGTGCACGTCAAGAAGTCTGACAGGGGTCGACTGCAAACTGAAGATCCTATGGAGATGG 2919
Db      71717 TTTTTCATATAAGAAAGGACGTCGGGGCGAGGATGCGAAACACGGCCGGACCCGAGGATG 71658

Qy       2920 ATAAGGAGAATGATATGTATGCTG--------------------------GGTATCATG 2952
Db      71657 AAGAGGCTGCAAGAAACTCAGCTGCTTCAACGAAGTCATCTGGACCGGTGAAGAATCAAG 71598
```

```
Qy       2953 AAAGTGATACTTCAGAAGCTGCTAATGAAGCAAAGATTCAAAAACTTAAAGAGCTGTCCA 3012

Db      71597 GTCCTGGACAGCAAATGGAGTTCACTCCGGAGCGCATGGCCGAGATAGAAGAGCTCGGTC 71538

Qy       3013 AGCTCCCGGGCATTTATGATAGACTTTCAAGGTCGCTGGCTCCAAGCATTTGGGAGCTTG 3072

Db      71537 GTTCGCCGGACATTTATGACCGGCTTGTGGCGTCGTTGGCCCCGTCAATTTGGGAGTTAG 71478

Qy       3073 AAGATATTAAAAAGGGTCTTCTTTGCCAGCTCTTTGGTGGGAAGGCTAAGAAAATTCCAT 3132

Db      71477 AAGACGTTAAAAAAGGACTGTTGTGTCAGCTCTTTGGCGCCACGAACAAGAGCTTCAGCG 71418

Qy       3133 CTGGAGCATC------TTTCCGAGGTGACATCAATGTTTTACTTGTTGGGGACCCTGGTA 3186

Db      71417 ACAAGGCAGCGAACAAGGTCCGGGGCGACATCAACATTTTACTCGTCGGTGATCCCGGGG 71358

Qy       3187 CCAGTAAATCTCAGCTGCTTCAGTATGTGCACAAGATAGCTCCTCGTGGAATCTACACTA 3246

Db      71357 TCGCCAAGTCTCAGCTATTGACGTACGTTCATCGTATCGCGCCGCGTGGTATGTACACCT 71298

Qy       3247 GTGGGCGAGGAAGTTCGGCGGTTGGGCTGACAGCGTATGTAACGAAGGATCCAGAAACTC 3306

Db      71297 CAGGTAGAGGCTCGTCTGCGGTTGGTCTTACAGCGTACGTCACTCGCGATCCGGAGTCAA 71238

Qy       3307 GAGAGACGGTATTGGAGAGCGGAGCTTTGGTTCTTAGTGATCGTGGGATATGCTGTATCG 3366

Db      71237 AAGACATGGTGCTCGAGTCAGGAGCGCTCGTGCTTTCTGATCGCGGCATTTGTTGCATTG 71178

Qy       3367 ATGAGTTCGACAAAATGTCTGATAATGCCCGAAGCATGCTTCATGAGGTAATGGAGCAAC 3426

Db      71177 ATGAGTTCGACAAGATGTCTGACAGTGCTCGATCAATGTTGCACGAAGTGATGGAGCAGC 71118

Qy       3427 AAACGGTATCTGTAGCCAAAGGGGGTATCATTGCCTCGCTGAACGCTCGGACGTCTGTCC 3486

Db      71117 AGACGGTTTCCATCGCGAAGGCTGGAATCATTGCGGTGTTAAATGCACGGACTTCGGTCC 71058

Qy       3487 TTGCATGTGCAAATCCTAGTGGGTCCCGATACAATGCGCGCCTTTCTGTGATTGATAACA 3546

Db      71057 TTGCGTCAGCGAACCCAGTCGGAAGTCGTTACAACCCGAACATGTCCATGGTGGAGAACA 70998

Qy       3547 TCCAGCTTCCTCCAACTCTACTTTCTAGATTTGATTTAATTTACTTAATGCTCGACAAAC 3606

Db      70997 TTCAACTCCCGCCGACACTGTTGTCCCGTTTTGATTTGCTCTACCTATTGCTTGACCGTC 70938

Qy       3607 CAGACGAGCAAAACGATCGTCGTCTCGCCAGGCATCTCGTGGCTTTACACTATGAAAACT 3666

Db      70937 CTAATCCCGAGACGGATCGCCGTCTGGCGAGACACTTAGTCAGTTTGCACTACAAGAATC 70878

Qy       3667 ATGAAGTTTCAAAGCAGGACGCCTTAGATCTACAAACACTTACCGCGTATATCACCTATG 3726

Db      70877 CCCCGCAGAAGAAACGGGGCGTGATTTCTGCCGACTTGCTCACCGAGTACGTCTCGTATG 70818

Qy       3727 CTCGTCAGCATGTACATCCTACATTAAGTGATGAAGCTGCTGAAGATTTGATTAATGGCT 3786

Db      70817 CTCGCGCCAATGTACAACCGGTGCTCAGCGACGAAGCCTCGGAAGAGCTCGTGGAGGGAT 70758

Qy       3787 ATGTTGAGATGCGCCAAAAGGGCAACTTTCCTGGAAGCAGTAAAAAGGTGATAACAGCCA 3846

Db      70757 ATGTTGAGATGCGACGCATG------------GGTGGTAGTCGAAAAGTTATTACTGCAA 70710

Qy       3847 CACCTCGGCAACTCGAAAGTATGATTCGTATCAGTGAAGCCCTAGCTCGAATGAGATTTT 3906

Db      70709 CGCCGAGACAACTTGAGTCTCTCATTCGTCTCTCAGAGTCTCTAGCGCGCATGCGTTTAA 70650

Qy       3907 CTGAAGTGGTAGAGAAAGTTGATGCAGCAGAAGCTGTGCGCCTTTTAGACGTCGCTTTGC 3966

Db      70649 GTGCGGTGGTGGATCGTGACGACGCTAAGGAGGCGCTGCGATTGATGCGGGTGGCGATGC 70590

Qy       3967 AGCAATCTGCTACTGATCATGCAACAGGTACGATAGACATGGATCTTATCACGACTGGAG 4026

Db      70589 AACAATCCGCCGTTGATCCTCGCACTGGGACGATTGATATGGATAAGATATTGACCGGAC 70530

Qy       4027 TGTCGGCCAGCGAGCGT 4043

Db      70529 ATAGCGCCAGCGACCGT 70513
```

```
RESULT 9
SCU14731
LOCUS       SCU14731                5392 bp    DNA     linear   PLN 31-JAN-1996
DEFINITION  Saccharomyces cerevisiae Cdc54p (CDC54) gene, complete cds.
ACCESSION   U14731
VERSION     U14731.1  GI:608170
KEYWORDS    .
SOURCE      Saccharomyces cerevisiae (baker's yeast)
  ORGANISM  Saccharomyces cerevisiae
            Eukaryota; Fungi; Ascomycota; Saccharomycotina; Saccharomycetes;
            Saccharomycetales; Saccharomycetaceae; Saccharomyces.
REFERENCE   1  (bases 1 to 5392)
  AUTHORS   Whitebread,L.A. and Dalton,S.
  TITLE     Cdc54 belongs to the Cdc46/Mcm3 family of proteins which are
            essential for initiation of eukaryotic DNA replication
  JOURNAL   Gene 155 (1), 113-117 (1995)
   PUBMED   7698653
REFERENCE   2  (bases 1 to 5392)
  AUTHORS   Dalton,S.
  TITLE     Direct Submission
  JOURNAL   Submitted (15-SEP-1994) Stephen Dalton, Department of Gene
            Regulation, Roche Institute of Molecular Biology, 340 Kingsland
            Street, Nutley, NJ 07110-1199, USA
FEATURES             Location/Qualifiers
     source          1..5392
                     /organism="Saccharomyces cerevisiae"
                     /mol_type="genomic DNA"
                     /db_xref="taxon:4932"
     gene            1921..4722
                     /gene="CDC54"
     CDS             1921..4722
                     /gene="CDC54"
                     /codon_start=1
                     /product="Cdc54p"
                     /protein_id="AAA86310.1"
                     /db_xref="GI:608171"
                     /translation="MSQQSSSPTKEDNNSSSPVVPNPDSVPPQLSSPALFYSSSSSQG
                     DIYGRNNSQNLSQGEGNIRAAIGSSPLNFPSSSQRQNSDVFQSQGRQGRIRSSASASG
                     RSRYHSDLRSDRALPTSSSSLGRNGQNRVHMRRNDIHTSDLSSPRRIVDFDTRSGVNT
                     LDTSSSSAPPSEASEPLRIIWGTNVSIQECTTNFRNFLMSFKYKFRKILDEREEFINN
                     TTDEELYYIKQLNEMRELGTSNLNLDARNLLAYKQTEDLYHQLLNYPQEVISIMDQTI
                     KDCMVSLIVDNNLDYDLDEIETKFYKVRPYNVGSCKGMRELNPNDIDKLINLKGLVLR
                     STPVIPDMKVAFFKCNVCDHTMAVEIDRGVIQEPARCERIDCNEPNSMSLIHNRCSFA
                     DKQVIKLQETPDFVPDGQTPHSISLCVYDELVDSCRAGDRIEVTGTFRSIPIRANSRQ
                     RVLKSLYKTYVDVVHVKKVSDKRLDVDTSTIEQELMQNKVDHNEVEEVRQITDQDLAK
                     IREVAAREDLYSLLARSIAPSIYELEDVKKGILLQLFGGTNKTFTKGGRYRGDINILL
                     CGDPSTSKSQILQYVHKITPRGVYTSGKGSSAVGLTAYITRDVDTKQLVLESGALVLS
                     DGGVCCIDEFDKMSDSTRSVLHEVMEQQTISIAKAGIITTLNARSSILASANPIGSRY
                     NPNLPVTENIDLPPPLLSRFDLVYLVLDKVDEKNDRELAKHLTNLYLEDKPEHISQDD
                     VLPVEFLTMYISYAKEHIHPIITEAAKTELVRAYVGMRKMGDDSRSDEKRITATTRQL
                     ESMIRLAEAHAKMKLKNVVELEDVQEAVRLIRSAIKDYATDPKTGKIDMNLVQTGKSV
                     IQRKLQEDLSREIMNVLKDQASDSMSFNELIKQINEHSQDRVESSDIQEALSRLQQED
                     KVIVLGEGVRRSVRLNNRV"
ORIGIN

  Query Match             10.8%;  Score 470.8;  DB 15;  Length 5392;
  Best Local Similarity   55.6%;  Pred. No. 1.2e-128;
  Matches  948;  Conservative    0;  Mismatches  747;  Indels    9;  Gaps    2;

Qy       2364 AAGCATATTGAGGCCAGACCTTTCAATCTCAAAGCATCGGTGCACATGCGTGAACTCAAC 2423

Db       2815 AAGTTTTATAAAGTAAGGCCTTACAATGTAGGTTCTTGTAAAGGTATGCGCGAATTGAAT 2874

Qy       2424 CCTTCAGATATAGACAAATTGGTTTCTGTTAAAGGAATGGTTATCCGGTGCAGTTCTATC 2483

Db       2875 CCAAATGATATTGATAAATTGATAAACTTAAAGGGTCTTGTCCTGAGATCAACTCCGGTT 2934

Qy       2484 ATACCTGAAATTAAGGGGGCCTTCTTCAATGTTTAGTGTGTGGTCACTCGCCTCCGCTA 2543
```

```
Db   2935 ATCCCTGATATGAAGGTAGCGTTTTTCAAATGCAACGTCTGCGATCA---TACAATGGCA 2991

Qy   2544 GTTACAGTTGTTAAAGGGCGGGTTGAGGAGCCAACAAGGTGTGAAAAGCCAGAATGTGCA 2603

Db   2992 GTGGAAATTGATAGAGGAGTTATACAAGAGCCCGCTAGGTGTGAACGTATTGATTGTAAT 3051

Qy   2604 GCACGGAATGCTATGTCTCTTATTCACAATCGATGCACTTTTGCAAATAAGCAGATAGTG 2663

Db   3052 GAACCCAATTCCATGTCACTGATTCACAACAGGTGTTCATTCGCAGATAAACAAGTCATT 3111

Qy   2664 CGTCTTCAAGAAACTCCAGATGCCATTCCTGAAGGAGAGACTCCACACACAGTCAGCATG 2723

Db   3112 AAGTTACAGGAAACTCCAGACTTTGTGCCTGATGGACAAACGCCTCACTCTATCTCATTA 3171

Qy   2724 TGTTTATACAACACTATGGTTGATGCTGTGAAGCCTGGAGATCGTATTGAGGTAACAGGA 2783

Db   3172 TGTGTTTACGATGAATTAGTGGATTCTTGCAGGGCGGGCGATCGTATTGAGGTGACTGGC 3231

Qy   2784 GTTTTCAAGGCCATGGCAGTTCGAGTTGGTCCGAATCAACGAACATTACGAGCATTGTAT 2843

Db   3232 ACGTTCAGGTCCATCCCCATTAGAGCTAATTCCAGGCAACGCGTACTAAAGTCGTTGTAT 3291

Qy   2844 AAGACCTACATCGATTGCGTGCACGTCAAGAAGTCTGACAGGGGTCGACTGCAAACTGAA 2903

Db   3292 AAAACATACGTCGATGTGGTCCACGTTAAAAAAGTTTCAGATAAGAGGTTAGACGTCGAT 3351

Qy   2904 GATCCTATGGAGATGGATAAGGAGAATGATATGTATGCTGGGTATCATGAAAGTGATACT 2963

Db   3352 ACTTCTACTATTGAACAAGAATTAATGCAGAACAAGGTAGATCATAACGAGGTCGAAGAA 3411

Qy   2964 TCAGAAGCTGCTAATGAAGCAAAGATTCAAAAACTTAAAGAGCTGTCCAAGCTCCCGGGC 3023

Db   3412 GTAAGACAAATTACTGATCAGGATTTAGCAAAAATTCGCGAGGTTGCGGCAAGAGAAGAT 3471

Qy   3024 ATTTATGATAGACTTTCAAGGTCGCTGGCTCCAAGCATTTGGGAGCTTGAAGATATTAAA 3083

Db   3472 TTATACAGTTTATTAGCCCGCTCTATTGCCCCAAGTATTTACGAGCTAGAAGATGTCAAG 3531

Qy   3084 AAGGGTCTTCTTTGCCAGCTCTTTGGTGGGAAGGCTAAGAAAATTCCATCTGGAGCATCT 3143

Db   3532 AAGGGTATATTACTTCAGCTATTTGGCGGCACGAATAAAACCTTTACCAAGGGTGGCCGT 3591

Qy   3144 TTCCGAGGTGACATCAATGTTTTACTTGTTGGGGACCCTGGTACCAGTAAATCTCAGCTG 3203

Db   3592 TATAGAGGTGACATAAATATTTTACTTTGTGGGGATCCTTCTACTTCCAAATCGCAAATT 3651

Qy   3204 CTTCAGTATGTGCACAAGATAGCTCCTCGTGGAATCTACACTAGTGGGCGAGGAAGTTCG 3263

Db   3652 TTGCAATACGTTCACAAAATTACTCCTCGTGGTGTGTATACTTCGGGTAAAGGTTCATCT 3711

Qy   3264 GCGGTTGGGCTGACAGCGTATGTAACGAAGGATCCAGAAACTCGAGAGACGGTATTGGAG 3323

Db   3712 GCCGTTGGTTTAACTGCTTATATTACAAGGGATGTCGACACAAAACAACTTGTTTTGGAA 3771

Qy   3324 AGCGGAGCTTTGGTTCTTAGTGATCGTGGGATATGCTGTATCGATGAGTTCGACAAAATG 3383

Db   3772 AGTGGTGCATTAGTATTGTCTGATGGAGGTGTTTGTTGTATTGACGAGTTTGATAAAATG 3831

Qy   3384 TCTGATAATGCCCGAAGCATGCTTCATGAGGTAATGGAGCAACAAACGGTATCTGTAGCC 3443

Db   3832 AGTGATTCTACAAGGTCCGTCTTGCACGAAGTCATGGAACAGCAGACTATTTCAATCGCA 3891

Qy   3444 AAAGGGGGTATCATTGCCTCGCTGAACGCTCGGACGTCTGTCCTTGCATGTGCAAATCCT 3503

Db   3892 AAAGCGGGAATTATCACAACACTTAATGCCAGAAGTTCTATTTTGGCCAGTGCTAACCCA 3951

Qy   3504 AGTGGGTCCCGATACAATGCGCGCCTTTCTGTGATTGATAACATCCAGCTTCCTCCAACT 3563

Db   3952 ATTGGTTCACGCTACAACCCCAATTTGCCTGTGACTGAAAATATTGATCTACCGCCCCCA 4011

Qy   3564 CTACTTTCTAGATTTGATTTAATTTACTTAATGCTCGACAAACCAGACGAGCAAAACGAT 3623

Db   4012 CTACTTTCGAGATTCGATCTGGTCTATCTTGTCCTTGATAAGGTTGATGAGAAAAATGAC 4071

Qy   3624 CGTCGTCTCGCCAGGCATCTCGTGGCTTTACACTATGAAAA------CTATGAAGTTTCA 3677

Db   4072 AGAGAACTAGCCAAACACTTAACAAATCTTTACCTGGAAGATAAGCCCGAACATATTTCT 4131

Qy   3678 AAGCAGGACGCCTTAGATCTACAAACACTTACCGCGTATATCACCTATGCTCGTCAGCAT 3737

Db   4132 CAAGACGACGTGCTACCAGTTGAATTTTTAACGATGTATATTAGTTATGCAAAGGAGCAC 4191

Qy   3738 GTACATCCTACATTAAGTGATGAAGCTGCTGAAGATTTGATTAATGGCTATGTTGAGATG 3797

Db   4192 ATACACCCAATAATCACCGAGGCCGCTAAGACTGAGCTTGTTCGTGCTTATGTAGGAATG 4251

Qy   3798 CGCCAAAAGGGCAACTTTCCTGGAAGCAGTAAAAAGGTGATAACAGCCACACCTCGGCAA 3857

Db   4252 AGAAAGATGGGTGACGATTCGAGATCCGATGAGAAGAGAATCACAGCTACCACAAGACAA 4311

Qy   3858 CTCGAAAGTATGATTCGTATCAGTGAAGCCCTAGCTCGAATGAGATTTTCTGAAGTGGTA 3917

Db   4312 CTTGAAAGTATGATTCGTTTGGCTGAGGCGCACGCCAAAATGAAATTGAAAAACGTCGTA 4371

Qy   3918 GAGAAAGTTGATGCAGCAGAAGCTGTGCGCCTTTTAGACGTCGCTTTGCAGCAATCTGCT 3977

Db   4372 GAGCTGGAGGATGTTCAAGAAGCCGTTAGATTAATTAGATCAGCCATAAAAGATTATGCA 4431

Qy   3978 ACTGATCATGCAACAGGTACGATAGACATGGATCTTATCACGACTGGAGTGTCGGCCAGC 4037

Db   4432 ACAGACCCTAAAACCGGTAAAATCGACATGAATTTAGTTCAAACAGGTAAATCAGTTATT 4491

Qy   4038 GAGCGTATTCGTCGGGCCAACTTG 4061

Db   4492 CAGAGAAAACTACAGGAGGATTTG 4515
```

RESULT 10
SC9531X

```
LOCUS       SC9531X              27166 bp    DNA     linear   PLN 18-APR-2005
DEFINITION  S.cerevisiae chromosome XVI cosmid 9531.
ACCESSION   Z49919 U00094
VERSION     Z49919.1  GI:887584
KEYWORDS    CDC21 homologue; DSS4; F1Fo-ATP synthase g subunit; Grave's
            disease; guanine-releasing protein; Mitochondrial carrier protein;
            RNA polymerase I; RPA2; transfer RNA-Lys; zinc finger protein.
SOURCE      Saccharomyces cerevisiae (baker's yeast)
  ORGANISM  Saccharomyces cerevisiae
            Eukaryota; Fungi; Ascomycota; Saccharomycotina; Saccharomycetes;
            Saccharomycetales; Saccharomycetaceae; Saccharomyces.
REFERENCE   1  (bases 1 to 27166)
  AUTHORS   Bowman,S.
  JOURNAL   Unpublished
REFERENCE   2  (bases 1 to 27166)
  AUTHORS   Barrell,B.G., Rajandream,M.A. and Walsh,S.V.
  TITLE     Direct Submission
  JOURNAL   Submitted (23-JUN-1995) Saccharomyces cerevisiae chromosome XVI
            sequencing project, Sanger Centre, Hinxton Hall, Hinxton, Cambridge
            CB10 1RQ E-mail: barrell@sanger.ac.uk
COMMENT     Notes:
            All CDS over 100 codons have been analysed. CDS that are completely
            overlapped and those that are overlapped by more than 50%
            of their length by a larger CDS have been omitted from this
            analysis.
            Details of the omitted CDS are available on request. The more
            significant matches with motifs in the PROSITE database are
            also included but some of these may be fortuitous. The length in
            codons and the calculated codon adaptation index (CAI)
            is given for each CDS.
            Cosmid 9531 is overlapped at the 5' end by cosmid 9723, embl entry
            SC9723, accession no. Z48951, and at the 3' end by cosmid 9367,
            embl entry SC9367, accession no. Z49274.
FEATURES             Location/Qualifiers
     source          1..27166
                     /organism="Saccharomyces cerevisiae"
                     /mol_type="genomic DNA"
```

```
                     /strain="AB972"
                     /db_xref="taxon:4932"
                     /chromosome="XVI"
                     /map="16R"
                     /clone="cosmid 9531"
     CDS             <1..1727
                     /note="YP9531.01, unknown, len: 574, CAI: 0.14, overlaps
                     and extends cosmid 9723 orf YP9723.08 which has similarity
                     at N-terminus to transcription factors"
                     /codon_start=3
                     /product="unknown"
                     /protein_id="CAA90152.1"
                     /db_xref="GI:887585"
                     /db_xref="GOA:Q12753"
                     /db_xref="UniProt/Swiss-Prot:Q12753"
                     /translation="SISRSQPMFERVLSSTSLDSNMLSGHGALSDTSSILTSTFLDSE
                     PGVGKISKDYHHVPSLASISSLQSSQSLDQNFSIPQSPPLSSMSFNFLTGNINETNQN
                     HSNHQHSKSGNNWQDSSVSLPAKADSRLNMMDKNNSVGLDLLGHSKRISPISNSRVGE
                     VSVPLEEYIPSDIDGVGRVTDKSSLVYDWPFDESIERNFSTTATAATGESKFDINDNC
                     NRINSKSYSKTNSMNGNGMNNSNNNNINSNGNDKNNNNSSRQEHQGNGLFDMFTDSSS
                     ISTLSRANLLLQEKIGSQENSVKQENYSKNPQLRHQLTSRSRSFIHHPANEYLKNTFG
                     NSHSNDIGKGVEVLSLTPSFMDIPEKERETERSPSSNYITDRPFTRKPRSSSIDVNHR
                     YPPMAPTTVATSPGALNNAVASNLDDQLSLTSLNSQPSSIANMMMDPSNLAEQSSIHS
                     VPQSINSPRMPKTGSRQDKNIHTKKEERNPLNNIHDLSQLENVPDEMNQMFSPPLKSM
                     NRPDAIRENSSSSNFIIQGNSMISTPSGRNDLPDTSPMSSIQTASPPSQLLTDQGFAD
                     LDNFMSSL"
     misc_feature    1..103
                     /note="overlap with cosmid 9723, positions 17558 to 17660;
                     EMBL:SC9723, Z48951"
     CDS             3177..3983
                     /note="YP9531.02 unknown, len: , CAI: , possible zinc
                     finger protein"
                     /codon_start=1
                     /product="unknown"
                     /protein_id="CAA90153.1"
                     /db_xref="GI:887586"
                     /db_xref="GOA:Q12286"
                     /db_xref="SGD:S000006213"
                     /db_xref="UniProt/TrEMBL:Q12286"
                     /translation="MKPNNRTCDVITNKDESLPALLLPALNSYTCDDSLLKGQISSNG
                     RYQPFGPSDCSLLPKRLNIQAGQGSMPVSSIQCADHSYSNWQKESEKTKLPKLGCPTE
                     YTEYYKTVSSGETTDSAVVSSIATNRLKRKRQRDGPSCDSCRIKKKIKCNATIIIFLQD
                     RNLISSISSNLHYTLSQDDINQFRMKFFRKLPDVMGTYEVIKHLDKIVLFKACTSCSR
                     RNQKNGKCLFSRGFTKSDMNVFPKINSKLKDKSIFEMTVDDYVAAGFQTL"
     gene            complement(4210..7821)
                     /gene="RPA2"
     CDS             complement(4210..7821)
                     /gene="RPA2"
                     /note="[9531.03c, RPA2 gene (RPA135, SRP3, RRN2)
                     SW:RPA2_YEAST P22138 DNA-directed RNA polymerase I 135 KD
                     polypeptide (EC 2.7.7.6) (A135)(RNA polymerase I subunit
                     2)"
                     /codon_start=1
                     /product="Rpa2p"
                     /protein_id="CAA90154.1"
                     /db_xref="GI:887587"
                     /db_xref="GOA:P22138"
                     /db_xref="InterPro:IPR007120"
                     /db_xref="InterPro:IPR007121"
                     /db_xref="InterPro:IPR007641"
                     /db_xref="InterPro:IPR007642"
                     /db_xref="InterPro:IPR007644"
                     /db_xref="InterPro:IPR007645"
                     /db_xref="InterPro:IPR007647"
                     /db_xref="InterPro:IPR009674"
                     /db_xref="SGD:S000006214"
                     /db_xref="UniProt/Swiss-Prot:P22138"
                     /translation="MSKVIKPPGQARTADFRTLERESRFINPPKDKSAFPLLQEAVQP
                     HIGSFNALTEGPDGGLLNLGVKDIGEKVIFDGKPLNSEDEISNSGYLGNKLSVSVEQV
                     SIAKPMSNDGVSSAVERKVYPSESRQRLTSYRGKLLLKLKWSVNNGEENLFEVRDCGG
                     LPVMLQSNRCHLNKMSPYELVQHKEESDEIGGYFIVNGIEKLIRMLIVQRRNHPMAII
                     RPSFANRGASYSHYGIQIRSVRPDQTSQTNVLHYLNDGQVTFRFSWRKNEYLVPVVMI
                     LKALCHTSDREIFDGIIGNDVKDSFLTDRLELLLRGFKKRYPHLQNRTQVLQYLGDKF
                     RVVFQASPDQSDLEVGQEVLDRIVLVHLGKDGSQDKFRMLLFMIRKLYSLVAGECSPD
                     NPDATQHQEVLLGGFLYGMILKEKIDEYLQNIIAQVRMDINRGMAINFKDKRYMSRVL
                     MRVNENIGSKMQYFLSTGNLVSQSGLDLQQVSGYTVVAEKINFYRFISHFRMVHRGSF
                     FAQLKTTTVRKLLPESWGFLCPVHTPDGSPCGLLNHFAHKCRISTQQSDVSRIPSILY
                     SLGVAPASHTFAAGPSLCCVQIDGKIIGWVSHEQGKIIADTLRYWKVEGKTPGLPIDL
                     EIGYVPPSTRGQYPGLYLFGGHSRMLRPVRYLPLDKEDIVGPFEQVYMNIAVTPQEIQ
                     NNVHTHVEFTPTNILSILANLTPFSDFNQSPRNMYQCQMGKQTMGTPGVALCHRSDNK
                     LYRLQTGQTPIVKANLYDDYGMDNFPNGFNAVVAVISYTGYDMDDAMIINKSADERGF
                     GYGTMYKTEKVDLALNRNRGDPITQHFGFGNDEWPKEWLEKLDEDGLPYIGTYVEEGD
                     PICAYFDDTLNKTKIKTYHSSEPAYIEEVNLIGDESNKFQELQTVSIKYRIRRTPQIG
                     DKFSSRHGQKGVCSRKWPTIDMPFSETGIQPDIIINPHAFPSRMTIGMFVESLAGKAG
                     ALHGIAQDSTPWIFNEDDTPADYFGEQLAKAGYNYHGNEPMYSGATGEELRADIYVGV
                     VYYQRLRHMVNDKFQVRSTGPVNSLTMQPVKGRKRHGGIRVGEMERDALIGHGTSFLL
                     QDRLLNSSDYTQASVCRECGSILTTQQSVPRIGSISTVCCRRCSMRFEDAKKLLTKSE
                     DGEKIFIDDSQIWEDGQGNKFVGGNETTTVAIPFVLKYLDSELSAMGIRLRYNVEPK"
     tRNA            8687..8759
                     /product="tRNA-Lys"
                     /note="tRNA-Lys gene, anticodon ctt, len: 73"
     CDS             complement(9687..10667)
                     /note="YP9531.04c, similar to SW:GDC_HUMAN P16260 Grave's
                     disease carrier protein (31.0% identity in 300 aa
                     overlap); contains 2 x PS00215 Mitochondrial energy
                     transfer proteins signature"
                     /codon_start=1
                     /product="unknown"
                     /protein_id="CAA90155.1"
                     /db_xref="GI:887588"
                     /db_xref="GOA:Q12251"
                     /db_xref="InterPro:IPR001993"
                     /db_xref="InterPro:IPR002067"
                     /db_xref="InterPro:IPR002113"
                     /db_xref="SGD:S000006215"
                     /db_xref="UniProt/TrEMBL:Q12251"
                     /translation="MAEVLTVLEQPNSIKDFLKQDSNIAFLAGGVAGAVSRTVVSPFE
                     RVKILLQVQSSTTSYNRGIFSSIRQVYHEEGTKGLFRGNGLNCIRIFPYSAVQFVVYE
                     ACKKKLFHVNGNNGQEQLTNTQRLFSGALCGGCSVVATYPLDLIKTRLSIQTANLSSL
                     NRSKAKSISKPPGIWQLLSETYRLEGGLRGLYRGVWPTSLGVVPYVALNFAVYEQLRE
                     FGVNSSDAQPSWKSNLYKLTIGAISGGVAQTITYPFDLLRRRFQVLAMGGNELGFRYT
                     SVWDALVTIGRAEGVSGYYKGLAANLFKVVPSTAVSWLVYEVVCDSVRNW"
     misc_feature    complement(10218..10244)
                     /note="PS00215 Mitochondrial energy transfer proteins
                     signature"
     misc_feature    complement(10518..10544)
                     /note="PS00215 Mitochondrial energy transfer proteins
                     signature"
     CDS             10934..11188
                     /note="YP9531.05, unknown, len: 84, CAI: 0.12"
                     /codon_start=1
                     /product="unknown"
                     /protein_id="CAA90156.1"
                     /db_xref="GI:887589"
                     /db_xref="SGD:S000006216"
                     /db_xref="UniProt/TrEMBL:Q04203"
                     /translation="MELSQYLNYAFSLAYYIIIHLLCLSYIYEIIHKHKNVFVRPSKL
                     EDALPLYKTGKNTNKEGDLRLISFIPLLLKSVKHSIRVYT"
     CDS             complement(11257..12210)
                     /note="YP9531.06c, unknown, len: 317, CAI: 0.13, putative
                     zinc finger protein, contains PS00028 Zinc finger, C2H2
                     type, domain"
                     /codon_start=1
                     /product="unknown"
                     /protein_id="CAA90157.1"
                     /db_xref="GI:887590"
                     /db_xref="GOA:Q12145"
                     /db_xref="InterPro:IPR007087"
                     /db_xref="SGD:S000006217"
                     /db_xref="UniProt/TrEMBL:Q12145"
                     /translation="MNFSLSKQSSEKQSSYTDKSRSPNIGMCTVNYKSNLPLQVSAVD
                     QLSTGKGTQFYGHNSYKNERECYNSTKINLPPISSLLPNFENNTPPNVDSRVQFPPQQ
                     VYQSMNVVPIVNEIYTPISMNATSDQYPIYYTESQQPIPHSQSPHLTSSAPLMMPVMV
                     PTVYKPLTPYDKEPITIASEPNFTAISMASHPNAALELCHDRPKSVPPGYGVLPTMQE
                     ASNGRTKSEPGAVLNGSATFSDWKTDTRISSTKLRKQCPVCGKICSRPSTLKTHYLIH
                     TGDTPFKCTWEGCTKSFNVKSNMLRHLKSHERKRNKVLNTT"
```

```
     misc_feature    complement(11293..11361)
                     /note="PS00028 Zinc finger, C2H2 type, domain"
     misc_feature    complement(11383..11445)
                     /note="PS00028 Zinc finger, C2H2 type, domain"
     CDS             complement(13814..14143)
                     /note="YP9531.07c unknown, len: 109, CAI: 0.08"
                     /codon_start=1
                     /product="unknown"
                     /protein_id="CAA90158.1"
                     /db_xref="GI:887591"
                     /db_xref="SGD:S000006218"
                     /translation="MSCIFSADLGVEYSCAESRITNLVLCILCIREEKAAPVVKKDKF
                     LFFIFLSSKENLFCEISSRSSVSEVKSLRAKNNFICPHAVIGFQGAKAQLGIRFPQLN
                     GLLRLQH"
     CDS             complement(16908..17651)
                     /note="YP9531.08c, unknown, len: 247, CAI: 0.11, putative
                     zinc finger protein, contains 2 x PS00028 Zinc finger,

  Query Match             10.8%;  Score 470.8;  DB 15;  Length 27166;
  Best Local Similarity   55.6%;  Pred. No. 1.5e-128;
  Matches  948;  Conservative    0;  Mismatches  747;  Indels    9;  Gaps    2;

Qy       2364 AAGCATATTGAGGCCAGACCTTTCAATCTCAAAGCATCGGTGCACATGCGTGAACTCAAC 2423
Db      24269 AAGTTTTATAAAGTAAGGCCTTACAATGTAGGTTCTTGTAAAGGTATGCGCGAATTGAAT 24328

Qy       2424 CCTTCAGATATAGACAAATTGGTTTCTGTTAAAGGAATGGTTATCCGGTGCAGTTCTATC 2483
Db      24329 CCAAATGATATTGATAAATTGATAAACTTAAAGGGTCTTGTCCTGAGATCAACTCCGGTT 24388

Qy       2484 ATACCTGAAATTAAGGGGGCCTTCTTCAAATGTTTAGTGTGTGGTCACTCGCCTCCGCTA 2543
Db      24389 ATCCCTGATATGAAGGTAGCGTTTTTCAAATGCAACGTCTGCGATCA---TACAATGGCA 24445

Qy       2544 GTTACAGTTGTTAAAGGGCGGGTTGAGGAGCCAACAAGGTGTGAAAAGCCAGAATGTGCA 2603
Db      24446 GTGGAAATTGATAGAGGAGTTATACAAGAGCCCGCTAGGTGTGAACGTATTGATTGTAAT 24505

Qy       2604 GCACGGAATGCTATGTCTCTTATTCACAATCGATGCACTTTTGCAAATAAGCAGATAGTG 2663
Db      24506 GAACCCAATTCCATGTCACTGATTCACAACAGGTGTTCATTCGCAGATAAACAAGTCATT 24565

Qy       2664 CGTCTTCAAGAAACTCCAGATGCCATTCCTGAAGGAGAGACTCCACACACAGTCAGCATG 2723
Db      24566 AAGTTACAGGAAACTCCAGACTTTGTGCCTGATGGACAAACGCCTCACTCTATCTCATTA 24625

Qy       2724 TGTTTATACAACACTATGGTTGATGCTGTGAAGCCTGGAGATCGTATTGAGGTAACAGGA 2783
Db      24626 TGTGTTTACGATGAATTAGTGGATTCTTGCAGGGCGGGCGATCGTATTGAGGTGACTGGC 24685

Qy       2784 GTTTTCAAGGCCATGGCAGTTCGAGTTGGTCCGAATCAACGAACATTACGAGCATTGTAT 2843
Db      24686 ACGTTCAGGTCCATCCCCATTAGAGCTAATTCCAGGCAACGCGTACTAAAGTCGTTGTAT 24745

Qy       2844 AAGACCTACATCGATTGCGTGCACGTCAAGAAGTCTGACAGGGGTCGACTGCAAACTGAA 2903
Db      24746 AAAACATACGTCGATGTGGTCCACGTTAAAAAAGTTTCAGATAAGAGGTTAGACGTCGAT 24805

Qy       2904 GATCCTATGGAGATGGATAAGGAGAATGATATGTATGCTGGGTATCATGAAAGTGATACT 2963
Db      24806 ACTTCTACTATTGAACAAGAATTAATGCAGAACAAGGTAGATCATAACGAGGTCGAAGAA 24865

Qy       2964 TCAGAAGCTGCTAATGAAGCAAAGATTCAAAAACTTAAAGAGCTGTCCAAGCTCCCGGGC 3023
Db      24866 GTAAGACAAATTACTGATCAGGATTTAGCAAAAATTCGCGAGGTTGCGGCAAGAGAAGAT 24925

Qy       3024 ATTTATGATAGACTTTCAAGGTCGCTGGCTCCAAGCATTTGGGAGCTTGAAGATATTAAA 3083
Db      24926 TTATACAGTTTATTAGCCCGCTCTATTGCCCCAAGTATTTACGAGCTAGAAGATGTCAAG 24985

Qy       3084 AAGGGTCTTCTTTGCCAGCTCTTTGGTGGGAAGGCTAAGAAAATTCCATCTGGAGCATCT 3143
Db      24986 AAGGGTATATTACTTCAGCTATTTGGCGGCACGAATAAAACCTTTACCAAGGGTGGCCGT 25045

Qy       3144 TTCCGAGGTGACATCAATGTTTTACTTGTTGGGGACCCTGGTACCAGTAAATCTCAGCTG 3203
Db      25046 TATAGAGGTGACATAAATATTTTACTTTGTGGGGATCCTTCTACTTCCAAATCGCAAATT 25105

Qy       3204 CTTCAGTATGTGCACAAGATAGCTCCTCGTGGAATCTACACTAGTGGGCGAGGAAGTTCG 3263
Db      25106 TTGCAATACGTTCACAAAATTACTCCTCGTGGTGTGTATACTTCGGGTAAAGGTTCATCT 25165

Qy       3264 GCGGTTGGGCTGACAGCGTATGTAACGAAGGATCCAGAAACTCGAGAGACGGTATTGGAG 3323
Db      25166 GCCGTTGGTTTAACTGCTTATATTACAAGGGATGTCGACACAAAACAACTTGTTTTGGAA 25225

Qy       3324 AGCGGAGCTTTGGTTCTTAGTGATCGTGGGATATGCTGTATCGATGAGTTCGACAAAATG 3383
Db      25226 AGTGGTGCATTAGTATTGTCTGATGGAGGTGTTTGTTGTATTGACGAGTTTGATAAAATG 25285

Qy       3384 TCTGATAATGCCCGAAGCATGCTTCATGAGGTAATGGAGCAACAAACGGTATCTGTAGCC 3443
Db      25286 AGTGATTCTACAAGGTCCGTCTTGCACGAAGTCATGGAACAGCAGACTATTTCAATCGCA 25345

Qy       3444 AAAGGGGGTATCATTGCCTCGCTGAACGCTCGGACGTCTGTCCTTGCATGTGCAAATCCT 3503
Db      25346 AAAGCGGGAATTATCACAACACTTAATGCCAGAAGTTCTATTTTGGCCAGTGCTAACCCA 25405

Qy       3504 AGTGGGTCCCGATACAATGCGCGCCTTTCTGTGATTGATAACATCCAGCTTCCTCCAACT 3563
Db      25406 ATTGGTTCACGCTACAACCCCAATTTGCCTGTGACTGAAAATATTGATCTACCGCCCCCA 25465

Qy       3564 CTACTTTCTAGATTTGATTTAATTTACTTAATGCTCGACAAACCAGACGAGCAAAACGAT 3623
Db      25466 CTACTTTCGAGATTCGATCTGGTCTATCTTGTCCTTGATAAGGTTGATGAGAAAAATGAC 25525

Qy       3624 CGTCGTCTCGCCAGGCATCTCGTGGCTTTACACTATGAAAA------CTATGAAGTTTCA 3677
Db      25526 AGAGAACTAGCCAAACACTTAACAAATCTTTACCTGGAAGATAAGCCCGAACATATTTCT 25585

Qy       3678 AAGCAGGACGCCTTAGATCTACAAACACTTACCGCGTATATCACCTATGCTCGTCAGCAT 3737
Db      25586 CAAGACGACGTGCTACCAGTTGAATTTTTAACGATGTATATTAGTTATGCAAAGGAGCAC 25645

Qy       3738 GTACATCCTACATTAAGTGATGAAGCTGCTGAAGATTTGATTAATGGCTATGTTGAGATG 3797
Db      25646 ATACACCCAATAATCACCGAGGCCGCTAAGACTGAGCTTGTTCGTGCTTATGTAGGAATG 25705

Qy       3798 CGCCAAAAGGGCAACTTTCCTGGAAGCAGTAAAAAGGTGATAACAGCCACACCTCGGCAA 3857
Db      25706 AGAAAGATGGGTGACGATTCGAGATCCGATGAGAAGAGAATCACAGCTACCACAAGACAA 25765

Qy       3858 CTCGAAAGTATGATTCGTATCAGTGAAGCCCTAGCTCGAATGAGATTTTCTGAAGTGGTA 3917
Db      25766 CTTGAAAGTATGATTCGTTTGGCTGAGGCGCACGCCAAAATGAAATTGAAAAACGTCGTA 25825

Qy       3918 GAGAAAGTTGATGCAGCAGAAGCTGTGCGCCTTTTAGACGTCGCTTTGCAGCAATCTGCT 3977
Db      25826 GAGCTGGAGGATGTTCAAGAAGCCGTTAGATTAATTAGATCAGCCATAAAAGATTATGCA 25885

Qy       3978 ACTGATCATGCAACAGGTACGATAGACATGGATCTTATCACGACTGGAGTGTCGGCCAGC 4037
Db      25886 ACAGACCCTAAAACCGGTAAAATCGACATGAATTTAGTTCAAACAGGTAAATCAGTTATT 25945

Qy       4038 GAGCGTATTCGTCGGGCCAACTTG 4061
Db      25946 CAGAGAAAACTACAGGAGGATTTG 25969


RESULT 11
SCCHRXVI
LOCUS       SCCHRXVI              165536 bp    DNA     linear   PLN 18-APR-2005
DEFINITION  S.cerevisiae chromosome XVI 165536 bp sequence, cen rightwards.
ACCESSION   Z71255
VERSION     Z71255.1  GI:1279666
KEYWORDS    ARO7; ATPase; beta-transducin; CCL1; CDC21 homologue; CDE1; CDE2;
```

```
            CDE3; centromere; CHL1; chorismate mutase; citrate synthase;
            coatomer zeta subunit; cyclin; delta element; DnaJ; DSS4; electron
            transport flavoprotein; eukaryotic iniation factor 5; F1Fo-ATP
            synthase g subunit; gamma adaptin; GLN1; glutamine synthetase;
            Grave's disease; guanine-releasing protein; HAL1; halotolerance;
            histidine tRNA synthetase; HTS1; isocitrate lyase; LPT1; LTR; MAK3;
            MFS1; Mitochondrial carrier protein; N-acetyltransferase;
            NADPH-cytochrome P450 reductase; negative regulator; NHP6A;
            nonhistone chromosomal protein 6A; OSD1; phenylalanyl-tRNA
            synthetase; phosphotyrosine protein phosphatase; polyposis;
            protamine P1; protein kinase; protein transport; regulatory
            protein; ribosomal protein L37a; RNA polymerase I; ROX1; RPA2;
            SEC8; SMK1; spermidine synthase; TIF5; TKL1; transfer RNA-Gly;
            transfer RNA-Lys; transfer RNA-Phe; transfer RNA-Ser;
            transketolase; tropomyosin; ubiquitin-activating enzyme; vacuolar
            H(+)-ATPase 54 kda subunit; VMA13; YMC1; YME1; YTA11; zinc finger
            protein.
SOURCE      Saccharomyces cerevisiae (baker's yeast)
  ORGANISM  Saccharomyces cerevisiae
            Eukaryota; Fungi; Ascomycota; Saccharomycotina; Saccharomycetes;
            Saccharomycetales; Saccharomycetaceae; Saccharomyces.
REFERENCE   1  (bases 1 to 165536)
  AUTHORS   Badcock,K., Bowman,S., Churcher,C.M., Pearson,D., Rajandream,M.A.,
            Walsh,S.V. and Barrell,B.G.
  TITLE     Direct Submission
  JOURNAL   Submitted (19-APR-1996) Saccharomyces cerevisiae chromosome IV
            sequencing project, Sanger Centre, Hinxton Hall, Hinxton, Cambridge
            CB10 1RQ. E-mail: barrell@sanger.ac.uk
COMMENT     SC8132X   Z48483      1 to 25540
            SC9723    Z48951    100 to 17660
            SC9531X   Z49919    104 to 27166
            SC9367    Z49274     10 to 42190
            SC3085    Z68111    790 to 10339   (PCR product)
            SC9499X   Z49219    134 to 43776
            The most significant matches using FASTA -o to a non-redundant
            database compiled from Swissprot 32 and PIR 46 and Wormpep9 are
            included.
            All CDS over 100 codons have been analysed. CDS that are largely or
            completely overlapped by a larger CDS have been omitted from this
            analysis.
            The more significant matches with motifs in the PROSITE database
            are also included but some of these may be fortuitous. The length
            in codons/amino-acids is given for each CDS as is the calculated
            codon adaptation index (CAI).
            tRNA genes and LTRs are also included.
            Notes:
            This sequence has been compiled from the following submissions of
            individual contigs.
FEATURES             Location/Qualifiers
     source          1. .165536
                     /organism="Saccharomyces cerevisiae"
                     /mol_type="genomic DNA"
                     /strain="AB972"
                     /db_xref="taxon:4932"
                     /chromosome="XVI"
     gene            <1. .3032
                     /gene="YPL012W"
     CDS             <1. .3032
                     /gene="YPL012W"
                     /note="YPL012W, unknown, incomplete, len: 1009, CAI: 0.19"
                     /codon_start=3
                     /product="unknown"
                     /protein_id="CAA95029.1"
                     /db_xref="GI:1314064"
                     /db_xref="GOA:Q7LH12"
                     /db_xref="InterPro:IPR000357"
                     /db_xref="UniProt/TrEMBL:Q7LH12"
                     /translation="SNLSNKQLKAQKTKEDINASVMRSLRLITSVVSTGQWPSSQIEP
                     LCDVLLGVTKSSEQYLVSASFECFESMFKTMAETTISSGLAENKYLRVLDTIFALKPS
                     NVDTLLTKSWIAVVIKGMSTYATHQPLKALRKIPGVFHIMCTYLASETPEVYQAASQC
                     LISILSESVKDDLLLYTPSVDEKVFKNVDEIISQIAKTFIDFLSIRYSHCSREILKIL
                     VAAFNKFRYRSNPHFLKSLKIVDTWRVNEEQFMDLRNEIELVIGASISAMGPEMILAE
                     APLNLDNPSSERPGRAWLLPLIRDYTKNANLATFQNELAPYIKSFQSKFDKVPEESIQ
                     LRVFQTIVDQIWSTLPRFCELPMDLRESFTDEFASELSSLLYSEVELRTTICHALKVL
                     AESNVSYAEESSSHNVLLLQRFPISEAQKNIEYLSTKSTNLLAVLFNVYTQTTPNARS
                     YILETIDQYLKITSKEDLEKTFNNVCGLLKNSMNEESSGNVNKEKKKPQLTATLLDLI
                     ICMITYLPVSSYSALFSMFSLTVNSADALIQKRAYRIITKLSELKSGSTAVAQFISDI
                     ENVMVDSASSVQTSAKAARLTAIKTIVELLPLDHLDFIVRTVAEVILSTKDVNEKSRE
                     TAFDTLICMGRKMNEPNGIIKLFQIPGYDPTTPDQSSSISEFFKIISAGLIGESQHMV
                     SSSITGYACLVFEFKNELDSGILMDIYDTIELYLTSNSREIVKSAIGFTKVCVLGLPE
                     ELMRPKVPELLLKLLRWSHEHTGHFKAKVKHIIERLIRRFGYDYIEANFPEEDRRLLT
                     NIRKMRNRNKRKDEEVTTGVSDVAATKGSRFMSAFDEAVYGSDEENDNGSDQEENVAG
                     GKMKNGAKQFIVESGDNPLDLLDSQTLAHISSTRPKKFNKNQNRARFNDDAFNFDSEG
                     KLVVKGQPKPSTNVDDPLSAVTSGINAYLEAVKSGPVRGQRNKLKFRKNGKDSDEFGD
                     DDDGEKDSRLMRGRVNQGNKIGKHNKKGPKFKSRKKL"
     gene            complement(3265. .4326)
                     /gene="YPL011C"
     CDS             complement(3265. .4326)
                     /gene="YPL011C"
                     /note="YPL011C, unknown, len: 353, CAI: 0.12"
                     /codon_start=1
                     /product="unknown"
                     /protein_id="CAA95030.1"
                     /db_xref="GI:1314065"
                     /db_xref="GOA:Q12297"
                     /db_xref="InterPro:IPR006565"
                     /db_xref="InterPro:IPR009072"
                     /db_xref="SGD:S000005932"
                     /db_xref="UniProt/Swiss-Prot:Q12297"
                     /translation="MTTNNDFYFALLRISILQLLKAQGFDRARPSLVDVMTDLYAKFL
                     SLLASEVSSIAQARCDQDDTIALQDITLALENLGIVKPTNVLDVYDENSELSSSRGME
                     KFKDWCIYSTQLTDARITALPTVELLQSEEKESDPLSAIPDYLNQLLQNKGAKQKLET
                     KNRKTELIEDLINNNGLDDWIKLVIARQRINMIERASKKESQNVPALPHIAGYKSSIL
                     SRHHHTTITNEDRMPSAMTPRDEDALTEIQENPFVTSKLPIMRKENRLENITLSFEDE
                     ELESLGEVEGPNQKSQENNNEESFKENNKSLTESPHGDDRDISMFQFDSNVDTKWAEQ
                     EDMDSTFQRRTSLDYGGYF"
     gene            4641. .5210
                     /gene="YPL010W"
     CDS             4641. .5210
                     /gene="YPL010W"
                     /note="YPL010W, unknown, len: 189, CAI: 0.19, 39.7%
                     identity in 156 aa overlap to coatomer zeta subunit
                     COPZ_BOVIN P35604"
                     /codon_start=1
                     /product="unknown"
                     /protein_id="CAA95031.1"
                     /db_xref="GI:1314066"
                     /db_xref="GOA:P53600"
                     /db_xref="InterPro:IPR000804"
                     /db_xref="InterPro:IPR011012"
                     /db_xref="SGD:S000005931"
                     /db_xref="UniProt/Swiss-Prot:P53600"
                     /translation="MSSLSLYTVQAVLILDQQGERIYAKYYQPPHRSDEGHQLLFNSV
                     KKQKEFEKQLYRKTHKQDSEILIFEDHLVLYKEYIDITIYLVASLEENEIVLQQGFSA
                     IRGALDLILNSGMDKKNIQENYDMVLLAIDETIDNGVILETDSNTIASRVSKPPTNEP
                     QMALDLDKGFLGAWGFAKSKFQERLQQGL"
     gene            complement(5443. .8559)
                     /gene="YPL009C"
     CDS             complement(5443. .8559)
                     /gene="YPL009C"
                     /note="YPL009C, unknown, len: 1038, CAI: 0.19, contains
                     glutamic acid rich region"
                     /codon_start=1
                     /product="unknown"
                     /protein_id="CAA95032.1"
                     /db_xref="GI:1314067"
                     /db_xref="GOA:Q12532"
                     /db_xref="InterPro:IPR008532"
                     /db_xref="SGD:S000005930"
                     /db_xref="UniProt/TrEMBL:Q12532"
                     /translation="MKQRISALDLLLLARELKQDLEGYRLSNIYNIADSSKQFLLKFN
                     KPDSKLNVVVDCGLRIYLTEFSRPIPPTPSGFVVKLRKHLKAKRLTALKQVDQDRILV
                     LQFADGHFYLVLEFFSAGNVILLDENRRIMALQRVVLEHENKVGQIYEMFDESLFTTN
                     NESADESIEKNRKAEYTSELVNEWIKAVQAKYESDITVIKQLNIQGKEGAKKKKVKVP
                     SIHKLLLSKVPHLSSDLLSKNLKVFNIDPSESCLNLLEETDSLAELLNSTQLEYNQLL
                     TTTDRKGYILAKRNENYISEKDTADLEFIYDTFHPFKPYINGGDTDSSCIIEVEGPYN
```

```
                     RTLDKFFSTIESSKYALRIQNQESQAQKKIDDARAENDRKIQALLDVQELNERKGHLI
                     IENAPLIEEVKLAVQGLIDQQMDWNTIEKLIKSEQKKGNRIAQLLNLPLNLKQNKISV
                     KLDLSSKELNTSSDEDNESEGNTTDSSSDSDSEDMESSKERSTKSMKRKSNEKINVTI
                     DLGLSAYANATEYFNIKKTSAQKQKKVEKNVGKAMKNIEVKIDQQLKKKLKDSHSVLK
                     KIRTPYFFEKYSWFISSEGFLVMMGKSPAETDQIYSKYIEDDDIYMSNSFNSHVWIKN
                     PEKTEVPPNTLMQAGILCMSSSEAWSKKISSSPWWCFAKNVSKFDGSDNSILPEGAFR
                     LKNENDQNHLPPAQLVMGFGFLWKVKTSGNEDNGDDDEEEEEEEEEEEEEEEEEEEEE
                     EEEKEEEEKEEEQQQDEDDSNEVNGLEKGGDSNDSTKNNSFEHDNLEKDIEKHCTISS
                     DTDSDSGNAKAKNDNSSTQRILDEPGVPISLIENINSNVRGKRGKLKKIQKKYADQDE
                     TERLLRLEALGTLKGIEKQQQRKKEEIMKREVREDRKNKREKQRRLQALKFTKKEKAR
                     VNYDKHKSELKPSLDKGDVVDDIIPVFAPWPALLKYKYKVKIQPGSAKKTKTLTEILH
                     YFKSRPLDGSSTDNEMDWPQEHEMIKGLKEQDLVLLLCVDKLKVTIAGQKSTKNGGNS
                     SKKGKKKR"
     gene            9008..11593
                     /gene="CHL1"
     CDS             9008..11593
                     /gene="CHL1"
                     /note="YPL008W, CHL1 gene, len: 861 CAI: 0.11
                     SW:CHL1_YEAST P22516, CHL1 protein, contains PS00690
                     DEAH-box subfamily ATP-dependent helicases signature"
                     /codon_start=1
                     /product="Chl1p"
                     /protein_id="CAA95033.1"
                     /db_xref="GI:1314068"
                     /db_xref="GOA:P22516"
                     /db_xref="InterPro:IPR001945"
                     /db_xref="InterPro:IPR002464"
                     /db_xref="InterPro:IPR006554"
                     /db_xref="InterPro:IPR006555"
                     /db_xref="InterPro:IPR010614"
                     /db_xref="SGD:S000005929"
                     /db_xref="UniProt/Swiss-Prot:P22516"
                     /translation="MDKKEYSETFYHPYKPYDIQVQLMETVYRVLSEGKKIAILESPT
                     GTGKTLSLICATMTWLRMNKADIFTRMETNIKTNEDDSENLSDDEPDWVIDTYRKSVL
                     QEKVDLLNDYEKHLNEINTTSCKQLKTMCDLDKEHGRYKSVDPLRKKRKGARHLDVSL
                     EEQDFIPRPYESDSENNDTSKSTRGGRISDKDYKLSELNSQIITLLDKIDGKVSRDPN
                     NGDRFDVTNQNPVKIYYASRTYSQLGQFTSQLRLPSFPSSFRDKVPDEKVKYLPLASK
                     KQLCINPKVMKWKTLEAINDACADLRHSKEGCIFYQNTNEWRHCPDTLALRDMIFSEI
                     QDIEDLVPLGKSLGICPYYASREALPIAEVVTLPYQYLLSESTRSSLQINLENSIVII
                     DEAHNLIETINSIYSSQISLEDLKNCHKGIVTYFNKFKSRLNPGNRVNLLKLNSLLMT
                     LIQFIVKNFKKIGQEIDPNDMFTGSNIDTLNIHKLLRYIKVSKIAYKIDTYNQALKEE
                     ESSKNENPIKETHKKSVSSQPLLFKVSQFLYCLTNLTSEGQFFFEKNYSIKYMLLEPS
                     KPFESILNQAKCVVLAGGTMEPMSEFLSNLLPEVPSEDITTLSCNHVIPKENLQTYIT
                     NQPELEFTFEKRMSPSLVNNHLFQFFVDLSKAVPKKGGIVAFFPSYQYLAHVIQCWKQ
                     NDRFATLNNVRKIFYEAKDGDDILSGYSDSVAEGRGSLLLAIVGGKLSEGINFQDDLC
                     RAVVMVGLPFPNIFSGELIVKRKHLAAKIMKSGGTEEEASRATKEFMENICMKAVNQS
                     VGRAIRHANDYANIYLLDVRYNRPNFRKKLSRWVQDSINSEHTTHQVISSTRKFFSMR
                     SLNSR"
     misc_feature    10169..10198

  Query Match             10.8%;  Score 470.8;  DB 15;  Length 165536;
  Best Local Similarity   55.6%;  Pred. No. 2e-128;
  Matches  948;  Conservative    0;  Mismatches  747;  Indels    9;  Gaps    2;

Qy       2364 AAGCATATTGAGGCCAGACCTTTCAATCTCAAAGCATCGGTGCACATGCGTGAACTCAAC 2423
Db      67267 AAGTTTTATAAAGTAAGGCCTTACAATGTAGGTTCTTGTAAAGGTATGCGCGAATTGAAT 67326

Qy       2424 CCTTCAGATATAGACAAATTGGTTTCTGTTAAAGGAATGGTTATCCGGTGCAGTTCTATC 2483
Db      67327 CCAAATGATATTGATAAATTGATAAACTTAAAGGGTCTTGTCCTGAGATCAACTCCGGTT 67386

Qy       2484 ATACCTGAAATTAAGGGGGCCTTCTTCAAATGTTTAGTGTGTGGTCACTCGCCTCCGCTA 2543
Db      67387 ATCCCTGATATGAAGGTAGCGTTTTTCAAATGCAACGTCTGCGATCA---TACAATGGCA 67443

Qy       2544 GTTACAGTTGTTAAAGGGCGGGTTGAGGAGCCAACAAGGTGTGAAAAGCCAGAATGTGCA 2603
Db      67444 GTGGAAATTGATAGAGGAGTTATACAAGAGCCCGCTAGGTGTGAACGTATTGATTGTAAT 67503

Qy       2604 GCACGGAATGCTATGTCTCTTATTCACAATCGATGCACTTTTGCAAATAAGCAGATAGTG 2663
Db      67504 GAACCCAATTCCATGTCACTGATTCACAACAGGTGTTCATTCGCAGATAAACAAGTCATT 67563

Qy       2664 CGTCTTCAAGAAACTCCAGATGCCATTCCTGAAGGAGAGACTCCACACACAGTCAGCATG 2723
Db      67564 AAGTTACAGGAAACTCCAGACTTTGTGCCTGATGGACAAACGCCTCACTCTATCTCATTA 67623

Qy       2724 TGTTTATACAACACTATGGTTGATGCTGTGAAGCCTGGAGATCGTATTGAGGTAACAGGA 2783
Db      67624 TGTGTTTACGATGAATTAGTGGATTCTTGCAGGGCGGGCGATCGTATTGAGGTGACTGGC 67683

Qy       2784 GTTTTCAAGGCCATGGCAGTTCGAGTTGGTCCGAATCAACGAACATTACGAGCATTGTAT 2843
Db      67684 ACGTTCAGGTCCATCCCCATTAGAGCTAATTCCAGGCAACGCGTACTAAAGTCGTTGTAT 67743

Qy       2844 AAGACCTACATCGATTGCGTGCACGTCAAGAAGTCTGACAGGGGTCGACTGCAAACTGAA 2903
Db      67744 AAAACATACGTCGATGTGGTCCACGTTAAAAAAGTTTCAGATAAGAGGTTAGACGTCGAT 67803

Qy       2904 GATCCTATGGAGATGGATAAGGAGAATGATATGTATGCTGGGTATCATGAAAGTGATACT 2963
Db      67804 ACTTCTACTATTGAACAAGAATTAATGCAGAACAAGGTAGATCATAACGAGGTCGAAGAA 67863

Qy       2964 TCAGAAGCTGCTAATGAAGCAAAGATTCAAAAACTTAAAGAGCTGTCCAAGCTCCCGGGC 3023
Db      67864 GTAAGACAAATTACTGATCAGGATTTAGCAAAAATTCGCGAGGTTGCGGCAAGAGAAGAT 67923

Qy       3024 ATTTATGATAGACTTTCAAGGTCGCTGGCTCCAAGCATTTGGGAGCTTGAAGATATTAAA 3083
Db      67924 TTATACAGTTTATTAGCCCGCTCTATTGCCCCAAGTATTTACGAGCTAGAAGATGTCAAG 67983

Qy       3084 AAGGGTCTTCTTTGCCAGCTCTTTGGTGGGAAGGCTAAGAAAATTCCATCTGGAGCATCT 3143
Db      67984 AAGGGTATATTACTTCAGCTATTTGGCGGCACGAATAAAACCTTTACCAAGGGTGGCCGT 68043

Qy       3144 TTCCGAGGTGACATCAATGTTTTACTTGTTGGGGACCCTGGTACCAGTAAATCTCAGCTG 3203
Db      68044 TATAGAGGTGACATAAATATTTTTACTTTGTGGGGATCCTTCTACTTCCAAATCGCAAATT 68103

Qy       3204 CTTCAGTATGTGCACAAGATAGCTCCTCGTGGAATCTACACTAGTGGGCGAGGAAGTTCG 3263
Db      68104 TTGCAATACGTTCACAAAATTACTCCTCGTGGTGTGTATACTTCGGGTAAAGGTTCATCT 68163

Qy       3264 GCGGTTGGGCTGACAGCGTATGTAACGAAGGATCCAGAAACTCGAGAGACGGTATTGGAG 3323
Db      68164 GCCGTTGGTTTAACTGCTTATATTACAAGGATGTCGACACAAAACAACTTGTTTTGGAA 68223

Qy       3324 AGCGGAGCTTTGGTTCTTAGTGATCGTGGGATATGCTGTATCGATGAGTTCGACAAAATG 3383
Db      68224 AGTGGTGCATTAGTATTGTCTGATGGAGGTGTTTGTTGTATTGACGAGTTTGATAAAATG 68283

Qy       3384 TCTGATAATGCCCGAAGCATGCTTCATGAGGTAATGGAGCAACAAACGGTATCTGTAGCC 3443
Db      68284 AGTGATTCTACAAGGTCCGTCTTGCACGAAGTCATGGAACAGCAGACTATTTCAATCGCA 68343

Qy       3444 AAAGGGGGTATCATTGCCTCGCTGAACGCTCGGACGTCTGTCCTTGCATGTGCAAATCCT 3503
Db      68344 AAAGCGGGAATTATCACAACACTTAATGCCAGAAGTTCTATTTTGGCCAGTGCTAACCCA 68403

Qy       3504 AGTGGGTCCCGATACAATGCGCGCCTTTCTGTGATTGATAACATCCAGCTTCCTCCAACT 3563
Db      68404 ATTGGTTCACGCTACAACCCCAATTTGCCTGTGACTGAAAATATTGATCTACCGCCCCCA 68463

Qy       3564 CTACTTTCTAGATTTGATTTAATTTACTTAATGCTCGACAAACCAGACGAGCAAAACGAT 3623
Db      68464 CTACTTTCGAGATTCGATCTGGTCTATCTTGTCCTTGATAAGGTTGATGAGAAAAATGAC 68523

Qy       3624 CGTCGTCTCGCCAGGCATCTCGTGGCTTTACACTATGAAAA------CTATGAAGTTTCA 3677
Db      68524 AGAGAACTAGCCAAACACTTAACAAATCTTTACCTGGAAGATAAGCCCGAACATATTTCT 68583

Qy       3678 AAGCAGGACGCCTTAGATCTACAAACACTTACCGCGTATATCACCTATGCTCGTCAGCAT 3737
Db      68584 CAAGACGACGTGCTACCAGTTGAATTTTTAACGATGTATATTAGTTATGCAAAGGAGCAC 68643
```

```
Qy       3738 GTACATCCTACATTAAGTGATGAAGCTGCTGAAGATTTGATTAATGGCTATGTTGAGATG 3797
Db      68644 ATACACCCAATAATCACCGAGGCCGCTAAGACTGAGCTTGTTCGTGCTTATGTAGGAATG 68703

Qy       3798 CGCCAAAAGGGCAACTTTCCTGGAAGCAGTAAAAAGGTGATAACAGCCACACCTCGGCAA 3857
Db      68704 AGAAAGATGGGTGACGATTCGAGATCCGATGAGAAGAGAATCACAGCTACCACAAGACAA 68763

Qy       3858 CTCGAAAGTATGATTCGTATCAGTGAAGCCCTAGCTCGAATGAGATTTTCTGAAGTGGTA 3917
Db      68764 CTTGAAAGTATGATTCGTTTGGCTGAGGCGCACGCCAAAATGAAATTGAAAAACGTCGTA 68823

Qy       3918 GAGAAAGTTGATGCAGCAGAAGCTGTGCGCCTTTTAGACGTCGCTTTGCAGCAATCTGCT 3977
Db      68824 GAGCTGGAGGATGTTCAAGAAGCCGTTAGATTAATTAGATCAGCCATAAAAGATTATGCA 68883

Qy       3978 ACTGATCATGCAACAGGTACGATAGACATGGATCTTATCACGACTGGAGTGTCGGCCAGC 4037
Db      68884 ACAGACCCTAAAACCGGTAAAATCGACATGAATTTAGTTCAAACAGGTAAATCAGTTATT 68943

Qy       4038 GAGCGTATTCGTCGGGCCAACTTG 4061
Db      68944 CAGAGAAAACTACAGGAGGATTTG 68967


RESULT 12
CR380957_02/c
WPCOMMENT
  Sequence split into 13 fragments  LOCUS CR380957 Accession CR380957
     Fragment Name        Begin       End
     CR380957_00              1    110000
     CR380957_01         100001    210000
     CR380957_02         200001    310000
     CR380957_03         300001    410000
     CR380957_04         400001    510000
     CR380957_05         500001    610000
     CR380957_06         600001    710000
     CR380957_07         700001    810000
     CR380957_08         800001    910000
     CR380957_09         900001   1010000
     CR380957_10        1000001   1110000
     CR380957_11        1100001   1210000
     CR380957_12        1200001   1302002
Continuation (3 of 13) of CR380957 from base 200001 (CR380957 Candida glabrata strain CB

  Query Match             10.7%;  Score 464.4;  DB 15;  Length 110000;
  Best Local Similarity   55.6%;  Pred. No. 1.5e-126;
  Matches  938;  Conservative    0;  Mismatches  741;  Indels    9;  Gaps    2;

Qy       2363 GAAGCATATTGAGGCCAGACCTTTCAATCTCAAAGCATCGGTGCACATGCGTGAACTCAA 2422
Db      20935 GAAGTTTTACAAAGTAAGACCTTACAATGTTGAAACTCAAAAAGGTATGAGAGAATTGAA 20876

Qy       2423 CCCTTCAGATATAGACAAATTGGTTTCTGTTAAAGGAATGGTTATCCGGTGCAGTTCTAT 2482
Db      20875 TCCAAATGATATAGATAAGTTGATAAGCCTGAAAGGGTTGGTTCTTAGAGCGACACCTGT 20816

Qy       2483 CATACCTGAAATTAAGGGGGCCTTCTTCAAATGTTTAGTGTGTGGTCACTCGCCTCCGCT 2542
Db      20815 TATACCAGACATGAAAGTAGCTTTTTTCAAATGTAATATTTGTGATCAC---ACCATGGC 20759

Qy       2543 AGTTACAGTTGTTAAAGGGCGGGTTGAGGAGCCAACAAGGTGTGAAAAGCCAGAATGTGC 2602
Db      20758 TGTTGAAATCGATCGTGGTGTAATTCAGGAACCAGCTAGATGTGAACGTGTTGATTGTAA 20699

Qy       2603 AGCACGGAATGCTATGTCTCTTATTCACAATCGATGCACTTTTGCAAATAAGCAGATAGT 2662
Db      20698 TGAGGCTAACTCTATGACGCTAATTCATAATAGATGCTCTTTCGCTGATAAACAAGTTAT 20639

Qy       2663 GCGTCTTCAAGAAACTCCAGATGCCATTCCTGAAGGAGAGACTCCACACACAGTCAGCAT 2722
Db      20638 TAAACTGCAAGAAACTCCTGATTTGGTTCCAGATGGACAAACACCTCACTCTGTTTCTCT 20579

Qy       2723 GTGTTTATACAACACTATGGTTGATGCTGTGAAGCCTGGAGATCGTATTGAGGTAACAGG 2782
Db      20578 ATGTGTTTACGACGAGTTAGTTGATAGTTGCCGTGCTGGTGACAGAATTGAAGTTACAGG 20519

Qy       2783 AGTTTTCAAGGCCATGGCAGTTCGAGTTGGTCCGAATCAACGAACATTACGAGCATTGTA 2842
Db      20518 TACCTTCAGATCAATCCCCATCAAGGCCAACTCTAGACAACGTGTCTTAAAATCTTTATA 20459

Qy       2843 TAAGACCTACATCGATTGCGTGCACGTCAAGAAGTCTGACAGGGGTCGACTGCAAACTGA 2902
Db      20458 CAAAACATACATTGATGTCGTCCATGTAAAAAAGGTCTCGAACACTAGAATTGGTGTTGA 20399

Qy       2903 AGATCCTATGGAGATGGATAAGGAGAATGATATGTATGCTGGGTATCATGAAAGTGATAC 2962
Db      20398 TGTTTCAACTATCGAACAAGAACTATTGCAAAATAAACTAGATAATAATGATGTCGAAGA 20339

Qy       2963 TTCAGAAGCTGCTAATGAAGCAAAGATTCAAAAACTTAAAGAGCTGTCCAAGCTCCCGGG 3022
Db      20338 AGTAAGACAAATCTCTGATGCTGAAATCGAGAAGATAAAGCAGGTTGCACAAAGACCCGA 20279

Qy       3023 CATTTATGATAGACTTTCAAGGTCGCTGGCTCCAAGCATTTGGGAGCTTGAAGATATTAA 3082
Db      20278 TCTTTATGATCTTCTTGCGAGATCTATAGCGCCAAGTATTTATGAATTAGATGACGTCAA 20219

Qy       3083 AAAGGGTCTTCTTTGCCAGCTCTTTGGTGGGAAGGCTAAGAAAATTCCATCTGGAGCATC 3142
Db      20218 AAAGGGTATATTACTACAATTATTTGGTGGTACTAACAAAACGTTTAAGAAGGGTGGTCG 20159

Qy       3143 TTTCCGAGGTGACATCAATGTTTTACTTGTTGGGGACCCTGGTACCAGTAAATCTCAGCT 3202
Db      20158 TTATAGAGGTGATATTAACATTTTGCTATGTGGTGACCCCTCCACATCCAAATCCCAGAT 20099

Qy       3203 GCTTCAGTATGTGCACAAGATAGCTCCTCGTGGAATCTACACTAGTGGGCGAGGAAGTTC 3262
Db      20098 CTTACAATATGTTCATAAAATTGCTCCTAGAGGTGTCTACACATCTGGTAAGGGTTCATC 20039

Qy       3263 GGCGGTTGGGCTGACAGCGTATGTAACGAAGGATCCAGAAACTCGAGAGACGGTATTGGA 3322
Db      20038 AGCAGTAGGTCTAACGGCATATGTCACAAGAGATGTCGATAGTAAACAATTAGTTTTAGA 19979

Qy       3323 GAGCGGAGCTTTGGTTCTTAGTGATCGTGGGATATGCTGTATCGATGAGTTCGACAAAAT 3382
Db      19978 GAGTGGTGCCTTAGTGCTTTCTGATGGTGGTATCTGTTGTATTGATGAATTCGATAAAAT 19919

Qy       3383 GTCTGATAATGCCCGAAGCATGCTTCATGAGGTAATGGAGCAACAAACGGTATCTGTAGC 3442
Db      19918 GAGTGAATCAACTAGATCTGTTCTTCACGAAGTGATGGAACAACAAACGATCTCCGTTGC 19859

Qy       3443 CAAAGGGGGTATCATTGCCTCGCTGAACGCTCGGACGTCTGTCCTTGCATGTGCAAATCC 3502
Db      19858 TAAAGCTGGTATTATCACAACATTGAATGCTAGAAGTTCCATTTAGCTAGTGCAAACCC 19799

Qy       3503 TAGTGGGTCCCGATACAATGCGCGCCTTTCTGTGATTGATAACATCCAGCTTCCTCCAAC 3562
Db      19798 TATTGGATCCCGTTATAACCCTAATCTACCTGTTACTGAAAATATTGATCTTCCACCCCC 19739

Qy       3563 TCTACTTTCTAGATTTGATTTAATTTACTTAATGCTCGACAAACCAGACGAGCAAAACGA 3622
Db      19738 ATTACTGTCAAGATTCGACCTTGTTTATATTATTTTGGATAAAGTTGATGAAAGCACTGA 19679

Qy       3623 TCGTCGTCTCGCCAGGCATCTCGTGGCTTTACACTATGAAAACTA------TGAAGTTTC 3676
Db      19678 CCGTGATCTGGCAAAACACTTGACCAGTCTTTATCTAGAGGACAAGCCTGCACATGTTAC 19619

Qy       3677 AAAGCAGGACGCCTTAGATCTACAAACACTTACCGCGTATATCACCTATGCTCGTCAGCA 3736
Db      19618 TACCGATGATGTTCTCCCAATCGATTTCTTAACACAATACATCAATTATGTCAAACAGAA 19559

Qy       3737 TGTACATCCTACATTAAGTGATGAAGCTGCTGAAGATTTGATTAATGGCTATGTTGAGAT 3796
Db      19558 TGTTCATCCACTTGTAACAGAGCAAGCAAAAAATGAACTTGTAAAGGCTTATGTTGGCAT 19499

Qy       3797 GCGCCAAAAGGGCAACTTTCCTGGAAGCAGTAAAAAGGTGATAACAGCCACACCTCGGCA 3856
```

```
Db      19498 GAGAAAAATGGGTGACGATTCAAGATCAGATGAAAAGAGAATTACTGCAACAACGAGACA 19439

Qy       3857 ACTCGAAAGTATGATTCGTATCAGTGAAGCCCTAGCTCGAATGAGATTTTCTGAAGTGGT 3916

Db      19438 GCTTGAAAGTATGATTCGTTTATCTGAAGCCCATGCCAAGATGAGATTATCAAGCACTGT 19379

Qy       3917 AGAGAAAGTTGATGCAGCAGAAGCTGTGCGCCTTTTAGACGTCGCTTTGCAGCAATCTGC 3976

Db      19378 AGATCTTGAAGACGTCCGTGAGGCTGTCAGATTAATGAAATCGGCTATTAAAGATTACGC 19319

Qy       3977 TACTGATCATGCAACAGGTACGATAGACATGGATCTTATCACGACTGGAGTGTCGGCCAG 4036

Db      19318 AACTGATCCAAAGACAGGTAAAATCGATATGAATTTGGTTCAAACTGGTAAATCAGTTAT 19259

Qy       4037 CGAGCGTA 4044

Db      19258 ACAGCGTA 19251
```

RESULT 13
CR382132_07
WPCOMMENT
Sequence split into 40 fragments LOCUS CR382132 Accession CR382132

Fragment Name	Begin	End
CR382132_00	1	110000
CR382132_01	100001	210000
CR382132_02	200001	310000
CR382132_03	300001	410000
CR382132_04	400001	510000
CR382132_05	500001	610000
CR382132_06	600001	710000
CR382132_07	700001	810000
CR382132_08	800001	910000
CR382132_09	900001	1010000
CR382132_10	1000001	1110000
CR382132_11	1100001	1210000
CR382132_12	1200001	1310000
CR382132_13	1300001	1410000
CR382132_14	1400001	1510000
CR382132_15	1500001	1610000
CR382132_16	1600001	1710000
CR382132_17	1700001	1810000
CR382132_18	1800001	1910000
CR382132_19	1900001	2010000
CR382132_20	2000001	2110000
CR382132_21	2100001	2210000
CR382132_22	2200001	2310000
CR382132_23	2300001	2410000
CR382132_24	2400001	2510000
CR382132_25	2500001	2610000
CR382132_26	2600001	2710000
CR382132_27	2700001	2810000
CR382132_28	2800001	2910000
CR382132_29	2900001	3010000
CR382132_30	3000001	3110000
CR382132_31	3100001	3210000
CR382132_32	3200001	3310000
CR382132_33	3300001	3410000
CR382132_34	3400001	3510000
CR382132_35	3500001	3610000
CR382132_36	3600001	3710000
CR382132_37	3700001	3810000
CR382132_38	3800001	3910000
CR382132_39	3900001	4003362

Continuation (8 of 40) of CR382132 from base 700001 (CR382132 Yarrowia lipolytica chromo

Query Match 10.5%; Score 458.6; DB 15; Length 110000;
Best Local Similarity 56.5%; Pred. No. 8.4e-125;
Matches 939; Conservative 0; Mismatches 704; Indels 18; Gaps 4;

```
Qy       2408 CATGCGTGAACTCAACCCTTCAGATATAGACAAATTGGTTTCTGTTAAAGGAATGGTTAT 2467

Db      43821 CATGAGAGAGCTCAATCCTTCAGACATTGACAATCTTGTGTGCGTCAAGGGACTGTCTTT 43880

Qy       2468 CCGGTGCAGTTCTATCATACCTGAAATTAAGGGGGCCTTCTTCAAATGTTTAGTGTGTGG 2527

Db      43881 GAGATCAACGCAGGTCATTCCTGACATGAAGGAGGCGGTTTTCAAGTGCTCGGTGTGCAA 43940

Qy       2528 TCACTCGCCTCCGCTAGTTACAGTTGTTAAAGGGCGGGTTGAGGAGCCAACAAGGTGTGA 2587

Db      43941 CCACACCGAACCG---GTCCAAATCGAAAGAGGTATTATCGCCGAACCTGTCGTTTGTCC 43997

Qy       2588 AAAGCCAGAATGTGCAGCACGGAATGCTATGTCTCTTATTCACAATCGATGCACTTTTGC 2647

Db      43998 TCGAACTGTGTGCCAGAGTAAGTACTCCATGCAAATTCTGCACAACCGGTCCATTTTCGG 44057

Qy       2648 AAATAAGCAGATAGTGCGTCTTCAAGAAACTCCAGATGCCATTCCTGAAGGAGAGACTCC 2707

Db      44058 CGACAAACAGGTTGTGCGACTCCAAGAGACACCGGACGTTGTTCCCGACGGCCAGACGCC 44117

Qy       2708 ACACACAGTCAGCATGTGTTTATACAACACTATGGTTGATGCTGTGAAGCCTGGAGATCG 2767

Db      44118 CCACACGGTGTCTCTCTGTGTCTATGACGAGCTTGTGGACGCCTGCAAGGCCGGAGATCG 44177

Qy       2768 TATTGAGGTAACAGGAGTTTTCAAGGCCATGGCAGTTCGAGTTGGTCCGAATCAACGAAC 2827

Db      44178 GGTCCAGGTGACGGGTATTTACCGATCTGTACCTACGCGAGTGAATCCCCGAAACAGAAC 44237

Qy       2828 ATTACGAGCATTGTATAAGACCTACATCGATTGCGTGCACGTCAAGAAGTCTGACAGGGG 2887

Db      44238 CATCAGATCGCTGTTCAAGACCTACATTGACGTTGTTCACGTGCAAAAGACCGACAAGAC 44297

Qy       2888 TCGACTGCAAACTGAAGATCCTATGGAGATGGATAAGGAGAATGATATGTATGCTGGGTA 2947

Db      44298 TCGTGTCCAGGTTGATACCGACATATCTGGAGCCAACAACGAGGACGCGGCTGAGAAGAA 44357

Qy       2948 TCATGAAAGTGATACTTCAGAAGCTGCTAATGAAGCAAAGATTCAAAAACTTAAAGAGCT 3007

Db      44358 TGACGTGGAGGAGACGCGAAAAATCACCGTCGACGAAATGGACCGAA---TCAAGGAGGT 44414

Qy       3008 GTCCAAGCTCCCGGGCATTTATGATAGACTTTCAAGGTCGCTGGCTCCAAGCATTTGGGA 3067

Db      44415 GTCTGAACGAGCCGATCTCTACGACCTCATGTCCCGGTCTCTGGCTCCGTCCATCTACGA 44474

Qy       3068 GCTTGAAGATATTAAAAAGGGTCTTCTTTGCCAGCTCTTTGGTGGGAAGGCTAAGAAAAT 3127

Db      44475 ACACGACGACGTTAAGAAGGGTATCTTACTGCAGCTGTTTGGAGGCACAAACAAGACGTT 44534

Qy       3128 TCC------ATCTGGAGCATCTTTCCGAGGTGACATCAATGTTTTACTTGTTGGGGACCC 3181

Db      44535 CACCAAGGGAGGTGGACCCAAGTACCGAGGCGATATCAACGTTCTTCTGTGCGGAGACCC 44594

Qy       3182 TGGTACCAGTAAATCTCAGCTGCTTCAGTATGTGCACAAGATAGCTCCTCGTGGAATCTA 3241

Db      44595 CTCCACATCGAAATCTCAGCTGCTTCAATATGTGCATAAGATTGCTCCGCGAGGAGTGTA 44654

Qy       3242 CACTAGTGGGCGAGGAAGTTCGGCGGTTGGGCTGACAGCGTATGTAACGAAGGATCCAGA 3301

Db      44655 CACTTCAGGCAAGGGCTCTTCGGCCGTGGGTCTGACAGCATATGTGGCACGCGATCCCGA 44714

Qy       3302 AACTCGAGAGACGGTATTGGAGAGCGGAGCTTTGGTTCTTAGTGATCGTGGGATATGCTG 3361

Db      44715 TACTCGACAGCCCGTGCTTGAATCTGGCGCTCTGGTTCTTTCGGATGGCGGAGTTTGTTG 44774

Qy       3362 TATCGATGAGTTCGACAAAATGTCTGATAATGCCCGAAGCATGCTTCATGAGGTAATGGA 3421

Db      44775 CATTGATGAGTTTGACAAGATGAGTGACGCCACCAGATCGGTGTTGCACGAAGTCATGGA 44834

Qy       3422 GCAACAAACGGTATCTGTAGCCAAAGGGGGTATCATTGCCTCGCTGAACGCTCGGACGTC 3481

Db      44835 GCAGCAGACGGTGTCGATTGCCAAAGCCGGCATTATCACCACCTTGAACGCACGAACGTC 44894

Qy       3482 TGTCCTTGCATGTGCAAATCCTAGTGGGTCCCGATACAATGCGCGCCTTTCTGTGATTGA 3541

Db      44895 CATCATTGCCTCAGCCAACCCCATTGACTCTCGGTACAATCCCGATTTGCCCGTGACAAA 44954
```

```
Qy       3542 TAACATCCAGCTTCCTCCAACTCTACTTTCTAGATTTGATTTAATTTACTTAATGCTCGA 3601
Db      44955 GAACATTGATCTTCCTCCTACACTTTTGTCTCGTTTTGATCTCGTCTATCTAATGCTTGA 45014

Qy       3602 CAAACCAGACGAGCAAAACGATCGTCGTCTCGCCAGGCATCTCG------TGGCTTTACA 3655
Db      45015 CCAAGTCGATGAACGAGTCGATCGTCATCTGGCTACTCATATCGCAAACATGTACCTTGA 45074

Qy       3656 CTATGAAAACTATGAAGTTTCAAAGCAGGACGCCTTAGATCTACAAACACTTACCGCGTA 3715
Db      45075 CGACAACACCGCCACTGCTGCCACTGGAGAGATCTTACCGGTGGAGTTTCTGACTCTGTA 45134

Qy       3716 TATCACCTATGCTCGTCAGCATGTACATCCTACATTAAGTGATGAAGCTGCTGAAGATTT 3775
Db      45135 CCTGTCGTACGCACGAGCCAATGTCTTCCCTCAAATCACCCCCGACGCCAAAGACAGACT 45194

Qy       3776 GATTAATGGCTATGTTGAGATGCGCCAAAAGGGCAACTTTCCTGGAAGCAGTAAAAAGGT 3835
Db      45195 TGTGGCGGCATATGTGGACATGCGAAAGCAGGGAGATGATCCTCGGTCTTCGGAAAAACG 45254

Qy       3836 GATAACAGCCACACCTCGGCAACTCGAAAGTATGATTCGTATCAGTGAAGCCCTAGCTCG 3895
Db      45255 AATCACAGCCACCACCAGACAGCTGGAGAGTATGATTCGGCTGTCAGAGGCACATGCCAA 45314

Qy       3896 AATGAGATTTTCTGAAGTGGTAGAGAAAGTTGATGCAGCAGAAGCTGTGCGCCTTTTAGA 3955
Db      45315 GATGCGTCTTTCGTCCACAGTCGAGGTTTCTGACGTCGACGAAGCCGTGCGACTTATCAG 45374

Qy       3956 CGTCGCTTTGCAGCAATCTGCTACTGATCATGCAACAGGTACGATAGACATGGATCTTAT 4015
Db      45375 AGCCGCCATCAAGGATTACGCCACCGATCCTGTTTCTGGACGAATCGACATGGATCTGAT 45434

Qy       4016 CACGACTGGAGTGTCGGCCAGCGAGCGTATTCGTCGGGCCA 4056
Db      45435 TCAGACAGGAACTTCTCTGGCCCAGCGAACCGCGCAGGCCA 45475
```

RESULT 14
AE016817_06
WPCOMMENT
Sequence split into 15 fragments LOCUS AE016817 Accession AE016817

Fragment Name	Begin	End
AE016817_00	1	110000
AE016817_01	100001	210000
AE016817_02	200001	310000
AE016817_03	300001	410000
AE016817_04	400001	510000
AE016817_05	500001	610000
AE016817_06	600001	710000
AE016817_07	700001	810000
AE016817_08	800001	910000
AE016817_09	900001	1010000
AE016817_10	1000001	1110000
AE016817_11	1100001	1210000
AE016817_12	1200001	1310000
AE016817_13	1300001	1410000
AE016817_14	1400001	1466886

Continuation (7 of 15) of AE016817 from base 600001 (AE016817 Ashbya gossypii (= Eremoth

Query Match 10.5%; Score 457.8; DB 15; Length 110000;
Best Local Similarity 54.9%; Pred. No. 1.5e-124;
Matches 975; Conservative 0; Mismatches 787; Indels 15; Gaps 3;

```
Qy       2374 AGGCCAGACCTTTCAATCTCAAAGCATCGGTGCACATGCGTGAACTCAACCCTTCAGATA 2433
Db      54430 AGATCAGGCCGTACAATTTGGATTCAGAGAGGGGCATGAGAGAGCTCAACCCGAACGACA 54489

Qy       2434 TAGACAAATTGGTTTCTGTTAAAGGAATGGTTATCCGGTGCAGTTCTATCATACCTGAAA 2493
Db      54490 TCGATAAACTAGTAAGCATAAAGGGTCTTGTTCTTCGTTCCACTCCTGTCATTCCAGATA 54549

Qy       2494 TTAAGGGGGCCTTCTTCAAATGTTTAGTGTGTGGTCACTCGCCTCCGCTAGTTACAGTTG 2553
Db      54550 TGAAATTGGCATTTTTCAAATGTAGCGTATGTGACCATACCACGGCGGTAG---AAATTG 54606

Qy       2554 TTAAAGGGCGGGTTGAGGAGCCAACAAGGTGTGAAAAGCCAGAATGTGCAGCACGGAATG 2613
Db      54607 ATCGTGGTATTATCCAGGAGCCGTTGCGGTGCCCGCGTGTTGCATGTAATCAGAGAAACT 54666

Qy       2614 CTATGTCTCTTATTCACAATCGATGCACTTTTGCAAATAAGCAGATAGTGCGTCTTCAAG 2673
Db      54667 CCATGTCGTTAATCCACAATAGATGCTCCTTTGCAGATAAGCAAGTTATAAAGTTACAAG 54726

Qy       2674 AAACTCCAGATGCCATTCCTGAAGGAGAGACTCCACACACAGTCAGCATGTGTTTATACA 2733
Db      54727 AAACCCCGGATCTTGTCCCAGATGGCCAGACTCCCCACTCCGTGTCATTGTGTATATATG 54786

Qy       2734 ACACTATGGTTGATGCTGTGAAGCCTGGAGATCGTATTGAGGTAACAGGAGTTTTCAAGG 2793
Db      54787 ATGAGTTGGTTGACAGCTGCAGGGCGGGTGACAGGATTGAAGTGACAGGGATCTTCAGGT 54846

Qy       2794 CCATGGCAGTTCGAGTTGGTCCGAATCAACGAACATTACGAGCATTGTATAAGACCTACA 2853
Db      54847 CTATCCCCATAAGAGCAAATCAGCGGCAGCGTGCCCTTAAATCCCTGTACAAGACCTACT 54906

Qy       2854 TCGATTGCGTGCACGTCAAGAAGTCTGACAGGGGTCGACTGCAAACTGAAGATCCTATGG 2913
Db      54907 TGGATGTTGTTCACGTTCGAAAAGTCAGCGCACGGCGCCTCGATATTGA------TACCT 54960

Qy       2914 AGATGGATAAGGAGAATGATATGTATGCTGGGTATCATGAAAGTGATACTTCAGAAGCTG 2973
Db      54961 CCACTGTTGAACAGCAAATCCTCCAAAACCAGATGGACAATGTGGAGGAACTACGGAAAG 55020

Qy       2974 CTAATGAAGCAAAGATTCAAAAACTTAAAGAGCTGTCCAAGCTCCCGGGCATTTATGATA 3033
Db      55021 TAACTGATGAAGATATCGCCAAGATCAATGCTGTTGCAGCACGCCCGGATGTTTACGAAG 55080

Qy       3034 GACTTTCAAGGTCGCTGGCTCCAAGCATTTGGGAGCTTGAAGATATTAAAAAGGGTCTTC 3093
Db      55081 TCTTGGCCAGATCCATTGCGCCAAGTATCTACGAACTTGATGACATAAAGAAAGGAATTC 55140

Qy       3094 TTTGCCAGCTCTTTGGTGGGAAGGCTAAGAAAATTCCATCTGGAGCATCTTTCCGAGGTG 3153
Db      55141 TCCTGCAGCTATTCGGCGGGACAAATAAAACATTTACAAAGGGTGGCAGGTATAGGGGTG 55200

Qy       3154 ACATCAATGTTTTACTTGTTGGGGACCCTGGTACCAGTAAATCTCAGCTGCTTCAGTATG 3213
Db      55201 ATATAAATATACTTTTGTGCGGTGATCCATCTACTTCGAAGTCCCAAATCCTGCAGTACG 55260

Qy       3214 TGCACAAGATAGCTCCTCGTGGAATCTACACTAGTGGGCGAGGAAGTTCGGCGGTTGGGC 3273
Db      55261 TCCACAAGATTGCTCCTCGTGGTGTTTACACATCTGGTAAAGGTTCCTCTGCCGTTGGTC 55320

Qy       3274 TGACAGCGTATGTAACGAAGGATCCAGAAACTCGAGAGACGGTATTGGAGAGCGGAGCTT 3333
Db      55321 TAACAGCCTATATTACAAGAGATGCCGACACCAAGCAATCGGTGCTGGAGAGTGGCGCAC 55380

Qy       3334 TGGTTCTTAGTGATCGTGGGATATGCTGTATCGATGAGTTCGACAAAATGTCTGATAATG 3393
Db      55381 TAGTCCTATCTGACGGCGGCGTGTGCTGCATCGATGAATTCGACAAGATGAGCGACTCGA 55440

Qy       3394 CCCGAAGCATGCTTCATGAGGTAATGGAGCAACAAACGGTATCTGTAGCCAAAGGGGGTA 3453
Db      55441 CCCGTTCTGTTTTGCACGAAGTGATGAAGCAGCAACCAATATCCGTGGCCAAGGCGGGTA 55500

Qy       3454 TCATTGCCTCGCTGAACGCTCGGACGTCTGTCCTTGCATGTGCAAATCCTAGTGGGTCCC 3513
Db      55501 TCATCACCACTCTAAATGCCAGAACGTCCATATTGGCAAGTGCGAACCCCATAGGTTCAC 55560

Qy       3514 GATACAATGCGCGCCTTTCTGTGATTGATAACATCCAGCTTCCTCCAACTCTACTTTCTA 3573
Db      55561 GATACAACACCAATCTCCCCGTGACTGAGAACATTGACCTGCCCCCTCCATTGCTGTCTA 55620

Qy       3574 GATTTGATTTAATTTACTTAATGCTCGACAAACCAGACGAGCAAAACGATCGTCGTCTCG 3633
```

```
Db      55621 GATTCGACTTGGTATACCTGGTGTTGGACAAGGTCTCCGAGTCCACCGATCGAGAATTAG 55680

Qy       3634 CCAGGCATCT------CGTGGCTTTACACTATGAAAACTATGAAGTTTCAAAGCAGGACG 3687
Db      55681 CCAAGCACCTGACCAGCCTCTACCTAGAGGACAAGCCCGCCCATGTGTCCGAGTCGGACA 55740

Qy       3688 CCTTAGATCTACAAACACTTACCGCGTATATCACCTATGCTCGTCAGCATGTACATCCTA 3747
Db      55741 TCCTTCCGGTGCACTTTCTCACAATGTATATCAACTATGCCAAGCAGCATATTCACCCGG 55800

Qy       3748 CATTAAGTGATGAAGCTGCTGAAGATTTGATTAATGGCTATGTTGAGATGCGCCAAAAGG 3807
Db      55801 TGATCACAGAAGGGGCCAAGACCGAGCTGGTGCGCGCCTATGTGAACATGCGCAGCATGG 55860

Qy       3808 GCAACTTTCCTGGAAGCAGTAAAAAGGTGATAACAGCCACACCTCGGCAACTCGAAAGTA 3867
Db      55861 GCGACGACTCGCGCGCAGATGAGAAGAGAATCACCGCCACTACTCGGCAGCTGGAAAGCA 55920

Qy       3868 TGATTCGTATCAGTGAAGCCCTAGCTCGAATGAGATTTTCTGAAGTGGTAGAGAAAGTTG 3927
Db      55921 TGATCAGACTGTCCGAAGCGCACGCTAAGGTGCGCCTTTCGCAGCAGGTTGAGGTCAGCG 55980

Qy       3928 ATGCAGCAGAAGCTGTGCGCCTTTTAGACGTCGCTTTGCAGCAATCTGCTACTGATCATG 3987
Db      55981 ACGTCCAGGAAGCAGTGAGACTGATCAAGTCTGCCATCAAGGACTACGCAATAGACCCCA 56040

Qy       3988 CAACAGGTACGATAGACATGGATCTTATCACGACTGGAGTGTCGGCCAGCGAGCGTATTC 4047
Db      56041 AGACAGGCAAGATCGACATGAACCTTATCCAAACCGGCAAGTCCGTGATCCAGCGCAAGC 56100

Qy       4048 GTCGGGCCAACTTGCTAGCTGCTCTGCGAGAGCTTATAGCAGATAAAATTTCACCTGGCA 4107
Db      56101 TCCAGGAGGACCTTGCCCGTGAAATTGTCCGCCTGCTAACCGAACGCTCTGCCGATGTCA 56160

Qy       4108 GCTCCTCTGGCTTGAAGACCAGTCAGCTTCTTGAGGA 4144
Db      56161 TCACCTACAATGAGCTAGCCAGATTGCTGAACGAGAA 56197
```

RESULT 15
CR382123_15/c
WPCOMMENT
  Sequence split into 18 fragments LOCUS CR382123 Accession CR382123

Fragment Name	Begin	End
CR382123_00	1	110000
CR382123_01	100001	210000
CR382123_02	200001	310000
CR382123_03	300001	410000
CR382123_04	400001	510000
CR382123_05	500001	610000
CR382123_06	600001	710000
CR382123_07	700001	810000
CR382123_08	800001	910000
CR382123_09	900001	1010000
CR382123_10	1000001	1110000
CR382123_11	1100001	1210000
CR382123_12	1200001	1310000
CR382123_13	1300001	1410000
CR382123_14	1400001	1510000
CR382123_15	1500001	1610000
CR382123_16	1600001	1710000
CR382123_17	1700001	1753957

Continuation (16 of 18) of CR382123 from base 1500001 (CR382123 Kluyveromyces lactis str

```
  Query Match             10.5%;  Score 456;  DB 15;  Length 110000;
  Best Local Similarity   55.8%;  Pred. No. 5.1e-124;
  Matches  936;  Conservative    0;  Mismatches  725;  Indels   15;  Gaps    3;

Qy       2363 GAAGCATATTGAGGCCAGACCTTTCAATCTCAAAGCATCGGTGCACATGCGTGAACTCAA 2422
Db      39609 GAAGTTGTTCAAAGTGAGACCATATAATGTGGAAACCAAAAAGGGTATGAGAGAGCTTAA 39550

Qy       2423 CCCTTCAGATATAGACAAATTGGTTTCTGTTAAAGGAATGGTTATCCGGTGCAGTTCTAT 2482
Db      39549 TCCTAATGATATTGATAAATTGGTGAGTATCAAAGGTTTAGTTCTAAGATCCACACCTAT 39490

Qy       2483 CATACCTGAAATTAAGGGGGCCTTCTTCAAATGTTTAGTGTGTGGTCACTCGCCTCCGCT 2542
Db      39489 TATCCCAGATATGAGTGTTGCATTTTTCAAATGTAACGTATGCAATCA---TACTGTAGA 39433

Qy       2543 AGTTACAGTTGTTAAAGGGCGGGTTGAGGAGCCAACAAGGTGTGAAAAGCCAGAATGTGC 2602
Db      39432 AGTTGAGATCGATCGTGGTATTATTCAAGAACCAGTGAGATGTCCCAGAGTGGTCTGTAA 39373

Qy       2603 AGCACGGAATGCTATGTCTCTTATTCACAATCGATGCACTTTTGCAAATAAGCAGATAGT 2662
Db      39372 TAGCCCAAACTCCATGGTTTTAGTTCACAATAGGTGCACTTTCCAAGACAGACAAGTGAT 39313

Qy       2663 GCGTCTTCAAGAAACTCCAGATGCCATTCCTGAAGGAGAGACTCCACACACAGTCAGCAT 2722
Db      39312 AAAGTTACAAGAAACGCCTGATCTAGTCCCTGATGGACAAACTCCACACTCTGTTTCCCT 39253

Qy       2723 GTGTTTATACAACACTATGGTTGATGCTGTGAAGCCTGGAGATCGTATTGAGGTAACAGG 2782
Db      39252 ATGTGTATATGATGAGCTGGTTGATTCGTGTCGTGCTGGTGATAGAATTGAGGTTTCAGG 39193

Qy       2783 AGTTTTCAAGGCCATGGCAGTTCGAGTTGGTCCGAATCAACGAACATTACGAGCATTGTA 2842
Db      39192 TATCTTTAGATCTATTCCAATCAGGTCGAATCCAAAGCAAAGAGCCCTCAAATCTCTATA 39133

Qy       2843 TAAGACCTACATCGATTGCGTGCACGTCAAGAAGTCTGACAGGGGTCGACTGCAAACTGA 2902
Db      39132 CAAGACATACATCGATGTGGTTCATATCCAAAAGGTGGCTAAAGATAGAGTCGGTGTGGA 39073

Qy       2903 AGATCCTATGGAGATGGATAAGGAGAATGATATGTATGCTGGGTATCATGAAAGTGATAC 2962
Db      39072 CACTTCTACTGTTGAACAACAACTCTTGCAGAACCAGATTGATAATGTGGAAGAGATTA- 39014

Qy       2963 TTCAGAAGCTGCTAATGAAGCAAAGATTCAAAAACTTAAAGAGCTGTCCAAGCTCCCGGG 3022
Db      39013 -----GAACTTTGAGTAGCGAAGACATTAGGAGGATCAAAGAAACTGCTCGCCGATCGGA 38959

Qy       3023 CATTTATGATAGACTTTCAAGGTCGCTGGCTCCAAGCATTTGGGAGCTTGAAGATATTAA 3082
Db      38958 TGTCTATGATGTTTTGTCACGTTCCATCGCACCAAGTATTTACGAGTTAGATGACGTCAA 38899

Qy       3083 AAAGGGTCTTCTTTGCCAGCTCTTTGGTGGAAGGCTAAGAAAATTCCATCTGGAGCATC 3142
Db      38898 GAAGGGAATCTTATTACAACTCTTTGGTGGTGCTAATAAAACTTTCAAGAAAGGTGGAAG 38839

Qy       3143 TTTCCGAGGTGACATCAATGTTTTACTTGTTGGGGACCCTGGTACCAGTAAATCTCAGCT 3202
Db      38838 ATATAGAGGTGACATTAATATTCTACTATGTGGTGATCCTTCGACCTCTAAATCTCAAAT 38779

Qy       3203 GCTTCAGTATGTGCACAAGATAGCTCCTCGTGGAATCTACACTAGTGGGCGAGGAAGTTC 3262
Db      38778 TTTACAATATGTTCATAAGATTGCTCCTCGTGGTGTTTACACTTCGGGTAAGGGTTCCTC 38719

Qy       3263 GGCGGTTGGGCTGACAGCGTATGTAACGAAGGATCCAGAAACTCGAGAGACGGTATTGGA 3322
Db      38718 TGCTGTTGGTTTAACTGCTTATGTTACAAGAGATGTGGATACTAAGCAACTGGTGTTGGA 38659

Qy       3323 GAGCGGAGCTTTGGTTCTTAGTGATCGTGGGATATGCTGTATCGATGAGTTCGACAAAAT 3382
Db      38658 AAGTGGTGCTTTAGTTCTCTCTGACGGTGGTGTGTGTTGTATTGATGAATTTGATAAGAT 38599

Qy       3383 GTCTGATAATGCCCGAAGCATGCTTCATGAGGTAATGGAGCAACAAACGGTATCTGTAGC 3442
Db      38598 GAACGATAATACTAGATCAGTATTACATGAAGTTATGGAACAACAAACAATCTCTATTGC 38539

Qy       3443 CAAAGGGGGTATCATTGCCTCGCTGAACGCTCGGACGTCTGTCCTTGCATGTGCAAATCC 3502
Db      38538 AAAAGCAGGTATTATTACAACTTTAAATGCTAGAACTTCAATTCTGGCTAGTGCCAATCC 38479

Qy       3503 TAGTGGGTCCCGATACAATGCGCGCCTTTCTGTGATTGATAACATCCAGCTTCCTCCAAC 3562
```

```
Db     38478 TATCAACTCTAGGTACAATCCTAATCTTCCTGTGACTGAAAATATTGATTTGCCGCCTCC 38419

Qy      3563 TCTACTTTCTAGATTTGATTTAATTTACTTAATGCTCGACAAACCAGACGAGCAAAACGA 3622

Db     38418 ATTATTATCAAGATTTGATCTCGTTTATCTAGTGCTTGACAAGGTCAACGAAGCGTCTGA 38359

Qy      3623 TCGTCGTCTCGCCAGGCATCTCGTGGCTTTACACTATGAAAAC------TATGAAGTTTC 3676

Db     38358 TCGTGAACTCGCTAAGCATTTAACCAGTTTATACCTGGAAGACAGACCGGATTCTGTATC 38299

Qy      3677 AAAGCAGGACGCCTTAGATCTACAAACACTTACCGCGTATATCACCTATGCTCGTCAGCA 3736

Db     38298 TCAAGGTGATATTCTACCAGTTGAATTCTTGACAGCCTACATCAACTATGCAAAACAAAA 38239

Qy      3737 TGTACATCCTACATTAAGTGATGAAGCTGCTGAAGATTTGATTAATGGCTATGTTGAGAT 3796

Db     38238 CATTCACCCGGTAATAACTGAGTCAGCCAAGACTGAACTTGTGAGGGCATATGTCGGTAT 38179

Qy      3797 GCGCCAAAAGGGCAACTTTCCTGGAAGCAGTAAAAAGGTGATAACAGCCACACCTCGGCA 3856

Db     38178 GAGAAAGATGGGTGATGACAGTAGATCAGATGAAAAAAGAATTACCGCAACAACAAGACA 38119

Qy      3857 ACTCGAAAGTATGATTCGTATCAGTGAAGCCCTAGCTCGAATGAGATTTTCTGAAGTGGT 3916

Db     38118 GTTAGAAAGTATGATTAGACTTTCAGAAGCTCATGCTAAGATGAGGTTAAGTGAAAGAGT 38059

Qy      3917 AGAGAAAGTTGATGCAGCAGAAGCTGTGCGCCTTTTAGACGTCGCTTTGCAGCAATCTGC 3976

Db     38058 TGAACTTGAAGATGTGGAAGAAGCAGTGAGACTGATAAAGTCTGCCATCAAAGATTACGC 37999

Qy      3977 TACTGATCATGCAACAGGTACGATAGACATGGATCTTATCACGACTGGAGTGTCGG 4032

Db     37998 TACGGATCCTAAAACTGGTAAAATTGATATGAACCTTGTACAAACTGGTAAATCGG 37943
```

Search completed: December 6, 2005, 06:11:55
Job time : 14095 secs

GenCore version 5.1.6

OM nucleic - nucleic search, using sw model

Run on: December 5, 2005, 23:36:15 ; Search time 1603 Seconds
(without alignments)
18077.407 Million cell updates/sec

Title: US-10-768-511-5
Perfect score: 4348
Sequence: 1 atggcgcgccgcactcacgt..........gcgcaattcccagagctcgc 4348

Scoring table: IDENTITY_NUC
Gapop 10.0 , Gapext 1.0

Searched: 4996997 seqs, 3332346308 residues

Total number of hits satisfying chosen parameters: 9993994

Minimum DB seq length: 0
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries

Database : N_Geneseq_21:*
1: geneseqn1980s:*
2: geneseqn1990s:*
3: geneseqn2000s:*
4: geneseqn2001as:*
5: geneseqn2001bs:*
6: geneseqn2002as:*
7: geneseqn2002bs:*
8: geneseqn2003as:*
9: geneseqn2003bs:*
10: geneseqn2003cs:*
11: geneseqn2003ds:*
12: geneseqn2004as:*
13: geneseqn2004bs:*
14: geneseqn2005s:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	% Query Match	Length	DB	ID	Description
1	4348	100.0	4348	6	AAI67613	Aai67613 Cell cycl
2	4348	100.0	4348	10	AAD60813	Aad60813 Physcomit
3	880.2	20.2	2841	13	ADX53995	Adx53995 Plant ful
4	673	15.5	723	6	AAI67610	Aai67610 Cell cycl
5	673	15.5	723	10	AAD60810	Aad60810 Physcomit
6	634.2	14.6	1571	13	ADX36902	Adx36902 Plant ful
7	492.8	11.3	2970	10	ADB69927	Adb69927 C. neofor
8	490	11.3	1723	12	ADJ39580	Adj39580 Plant cDN
9	480.2	11.0	2733	6	ABZ32193	Abz32193 Candida a
10	470.8	10.8	2802	13	ADT47760	Adt47760 Bacterial
11	451.6	10.4	2625	8	ABT19388	Abt19388 Aspergill
12	451.6	10.4	3072	8	ABT21208	Abt21208 Aspergill
13	430.8	9.9	3273	6	ABL65258	Abl65258 Lung canc
14	430.8	9.9	3273	6	ABL65666	Abl65666 Lung canc
15	430.8	9.9	3273	6	ABL65259	Abl65259 Lung canc
16	430.8	9.9	3273	12	ADO19849	Ado19849 Human PRO
17	430.8	9.9	3273	12	ADO19847	Ado19847 Human PRO
18	430.8	9.9	3273	13	ADR26059	Adr26059 Breast ca
19	430.8	9.9	3273	13	ADU05807	Adu05807 Novel bro
20	430.4	9.9	3394	6	AAS94968	Aas94968 Human DNA
21	429.2	9.9	3395	10	ADJ56481	Adj56481 Frog cDNA
22	427.6	9.8	2860	4	AAH17708	Aah17708 Human cDN
23	427.6	9.8	3248	13	ACN38801	Acn38801 Tumour-as
24	419	9.6	2842	3	AAF15678	Aaf15678 Human pro
25	390.2	9.0	1105	13	ADX12413	Adx12413 Plant ful
26	352	8.1	2881	13	ADS49058	Ads49058 Bacterial
27	337.2	7.8	2944	8	ABT18794	Abt18794 Aspergill
28	337.2	7.8	3362	8	ABT20610	Abt20610 Aspergill
29	337.2	7.8	4944	8	ABT18200	Abt18200 Aspergill
30	337.2	7.8	5362	8	ABT20014	Abt20014 Aspergill
31	328	7.5	2692	4	ABL02965	Abl02965 Drosophil
32	321.4	7.4	595	13	ACN61030	Acn61030 Cotton gy
33	258.8	6.0	5265	12	ADP83382	Adp83382 Breast sp
34	255.4	5.9	3178	13	ACN41452	Acn41452 Human dia
35	243	5.6	770	4	AAF74193	Aaf74193 DNA encod
36	228	5.2	1899	8	ABT18930	Abt18930 Aspergill
37	228	5.2	1913	8	ABT18336	Abt18336 Aspergill
38	228	5.2	2445	8	ABT20750	Abt20750 Aspergill
39	228	5.2	2608	8	ABT20152	Abt20152 Aspergill
40	228	5.2	3913	8	ABT17742	Abt17742 Aspergill
41	228	5.2	4608	8	ABT19556	Abt19556 Aspergill
42	226.6	5.2	1638	13	ADT47932	Adt47932 Bacterial
43	222.4	5.1	5065	4	ABL02964	Abl02964 Drosophil
44	220.8	5.1	2676	13	ADT48060	Adt48060 Bacterial
45	219.4	5.0	2280	13	ADT47813	Adt47813 Bacterial

ALIGNMENTS

RESULT 1
AAI67613
ID AAI67613 standard; cDNA; 4348 BP.
XX
AC AAI67613;
XX
DT 27-FEB-2002 (first entry)
XX
DE Cell cycle protein 2 (CC-2) encoding cDNA.
XX
KW Cell Cycle Stress-Related Protein; CCSRP; cell cycle protein; CC-1; CC-2;
KW CC-3; environmental stress; ss.
XX
OS Physcomitrella patens.
XX
PN WO200177354-A2.
XX
PD 18-OCT-2001.
XX
PF 06-APR-2001; 2001WO-US011294.
XX
PR 07-APR-2000; 2000US-0196001P.
XX
PA (BADI) BASF PLANT SCI GMBH.
XX
PI Costa E SilvaO, Bohnert HJ, Van Thielen N, Chen R;
PI Sarria-Millan R;
XX
DR WPI; 2002-049151/06.
DR P-PSDB; AAG66003.
XX
PT Novel Cell Cycle Stress-Related Protein useful for increasing tolerance
PT to environmental stress, is selected from Cell Cycle Proteins 1-3, or
PT their orthologs.
XX
PS Claim 4; Fig 2B; 90pp; English.
XX
CC The invention relates to a Cell Cycle Stress-Related Protein (CCSRP),
CC isolated from Physcomitrella patens, and selected from Cell Cycle (CC)-1
CC protein, CC-2 protein, CC-3 protein, or their orthologs. The CCSRP and
CC encoding nucleic acids are useful for increasing tolerance to
CC environmental stress selected from salinity, drought and temperature, in

```
CC   transgenic plants including monocot and dicot selected from maize, wheat,
CC   rye, oat, triticale, rice, barley, soybean, peanut, cotton, rapeseed,
CC   canola, manihot, pepper, sunflower, tagetes, solanaceous plants, potato,
CC   tobacco, eggplant, tomato, Vicia species, pea, alfalfa, coffee, cacao,
CC   tea, Salix species, oil palm, coconut, perennial grass and forage crops.
CC   The nucleic acid is also useful for identifying organisms e.g. P. patens
CC   in a mixed population of microorganisms. The nucleic acids are also
CC   useful for evolutionary and protein structural studies. The proteins and
CC   nucleic acids are useful as markers for specific regions of the genome.
CC   The present sequence represents the cDNA encoding P. patens CC-2 protein
XX
SQ   Sequence 4348 BP; 1120 A; 987 C; 1093 G; 1148 T; 0 U; 0 Other;

  Query Match             100.0%;  Score 4348;  DB 6;  Length 4348;
  Best Local Similarity   100.0%;  Pred. No. 0;
  Matches 4348;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy          1 ATGGCGCGCCGCACTCACGTGAGGAATTGCACCTCCTTGTTCTGCGACGGTTCCATTCTT 60
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db          1 ATGGCGCGCCGCACTCACGTGAGGAATTGCACCTCCTTGTTCTGCGACGGTTCCATTCTT 60

Qy         61 TTTGGTTTTTAGTTTGCAAATCTTGATCGTGGAGTTGAGAAAAAGGGCGGTTCGTTGTCT 120
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db         61 TTTGGTTTTTAGTTTGCAAATCTTGATCGTGGAGTTGAGAAAAAGGGCGGTTCGTTGTCT 120

Qy        121 TGAGGTGTTCTTGTTGATTGTTCGTCATGGAAAATAATGATGCACTTGACATTGGAGCCG 180
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        121 TGAGGTGTTCTTGTTGATTGTTCGTCATGGAAAATAATGATGCACTTGACATTGGAGCCG 180

Qy        181 TGTCGTCCCCATATCCTTCGCAATCTGAAGGAGTGTCTACGCCATTGCCGCAAGTAACAT 240
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        181 TGTCGTCCCCATATCCTTCGCAATCTGAAGGAGTGTCTACGCCATTGCCGCAAGTAACAT 240

Qy        241 CACCGAGCTTCGACAATGCAGCCTCACCCGTGGCCGGGCGGAGGGCCGTACGGCAGACCC 300
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        241 CACCGAGCTTCGACAATGCAGCCTCACCCGTGGCCGGGCGGAGGGCCGTACGGCAGACCC 300

Qy        301 CTACATCTGCAGTTCGAAGGAGAGGGAGAGAAACGGATTCCGCTCGTCGTAGGAGGAGTC 360
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        301 CTACATCTGCAGTTCGAAGGAGAGGGAGAGAAACGGATTCCGCTCGTCGTAGGAGGAGTC 360

Qy        361 GATCTCGCAGTTTAGGCAATTCTGTTTATAGTTCCCCTTACGATGCGGGGACTCCTGGAA 420
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        361 GATCTCGCAGTTTAGGCAATTCTGTTTATAGTTCCCCTTACGATGCGGGGACTCCTGGAA 420

Qy        421 CTCCTGGAACTCCAGTGGCTACTCCGGTTTACGCTACCCCAGTCGGTACACCTATGGGTA 480
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        421 CTCCTGGAACTCCAGTGGCTACTCCGGTTTACGCTACCCCAGTCGGTACACCTATGGGTA 480

Qy        481 CCCCATCGTTCCATCGTGGCACGCCACAGTACAAACAGCGCAGTGAGCTTGGTTCCCAGG 540
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        481 CCCCATCGTTCCATCGTGGCACGCCACAGTACAAACAGCGCAGTGAGCTTGGTTCCCAGG 540

Qy        541 GGAAGCCTCTACATCGGAGACGTCGATCTCAATCCAGAGAACCCGGGCATCGATCTCCTT 600
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        541 GGAAGCCTCTACATCGGAGACGTCGATCTCAATCCAGAGAACCCGGGCATCGATCTCCTT 600

Qy        601 CAAGGGAACCTAGTGCTGATGGGCGTCCCTCTGAATCTGCTGAGCCAGATGACACTTTGG 660
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        601 CAAGGGAACCTAGTGCTGATGGGCGTCCCTCTGAATCTGCTGAGCCAGATGACACTTTGG 660

Qy        661 GTGGAGAATATGCTTATGTTTGGGGGACGAATGTTAACATTCCAGATGTGCTTAGGGCGA 720
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        661 GTGGAGAATATGCTTATGTTTGGGGGACGAATGTTAACATTCCAGATGTGCTTAGGGCGA 720

Qy        721 TTCGTCGATTTCTCCACAATTATCGTTCGAGTGCTCATGATCTTAATTCCAAGTACATCC 780
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        721 TTCGTCGATTTCTCCACAATTATCGTTCGAGTGCTCATGATCTTAATTCCAAGTACATCC 780

Qy        781 AGATCATAGAGGAGACTGTGGAGCGTGAGGAGGATACTCTAAATATCGACATGTCAGACA 840
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        781 AGATCATAGAGGAGACTGTGGAGCGTGAGGAGGATACTCTAAATATCGACATGTCAGACA 840

Qy        841 TTTATGACCATGATCCTGATCTATACGCAAAAATTGTTCGATACCCACTCGACATCATCC 900
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        841 TTTATGACCATGATCCTGATCTATACGCAAAAATTGTTCGATACCCACTCGACATCATCC 900

Qy        901 CCCTGTTGGACACTGAGTGTCAGGAAGTTGCTACCTCTTTACTACCAACGTTTGAGAAGC 960
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        901 CCCTGTTGGACACTGAGTGTCAGGAAGTTGCTACCTCTTTACTACCAACGTTTGAGAAGC 960

Qy        961 ATATTGAGGCCAGACCTTTCAATCTCAAAGCATCGGTGCACATGCGTGAACTCAACCCTT 1020
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        961 ATATTGAGGCCAGACCTTTCAATCTCAAAGCATCGGTGCACATGCGTGAACTCAACCCTT 1020

Qy       1021 CAGATATAGACAAATTGGTTTCTGTTAAAGGAATGGTTATCCGGTGCAGTTCTATCATAC 1080
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1021 CAGATATAGACAAATTGGTTTCTGTTAAAGGAATGGTTATCCGGTGCAGTTCTATCATAC 1080

Qy       1081 CTGAAATTAAGGGGGCCTTCTTCAAATGTTTAGTGTGTGGTCACTCGCCTCCGCTAGTTA 1140
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1081 CTGAAATTAAGGGGGCCTTCTTCAAATGTTTAGTGTGTGGTCACTCGCCTCCGCTAGTTA 1140

Qy       1141 CAGTTGTTAAAGGGCGGGTTGAGGAGCCAACAAGGTGTGAAAAGCCAGAATGTGCAGCAC 1200
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1141 CAGTTGTTAAAGGGCGGGTTGAGGAGCCAACAAGGTGTGAAAAGCCAGAATGTGCAGCAC 1200

Qy       1201 GGAATGCTATGTCTCTTATTCACAATCGATGCACTTTTGCAAATAAGCAGATAGTGCGTC 1260
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1201 GGAATGCTATGTCTCTTATTCACAATCGATGCACTTTTGCAAATAAGCAGATAGTGCGTC 1260

Qy       1261 TTCAAGAAACTCCAGATGCCATTCCTGAAGGAGAGACTCCACACACAGTCAGCATGTGTT 1320
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1261 TTCAAGAAACTCCAGATGCCATTCCTGAAGGAGAGACTCCACACACAGTCAGCATGTGTT 1320

Qy       1321 TATACAACACTATGGTTGATGCTGTGAAGCCTGGAGATCGTATTGAGGTAACAGGAGTTT 1380
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1321 TATACAACACTATGGTTGATGCTGTGAAGCCTGGAGATCGTATTGAGGTAACAGGAGTTT 1380

Qy       1381 TCAAGGCCATGGCAGTTCGAGTTGGTCATGGCGCGCCGCACTCACGTGAGGAATTGCACC 1440
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1381 TCAAGGCCATGGCAGTTCGAGTTGGTCATGGCGCGCCGCACTCACGTGAGGAATTGCACC 1440

Qy       1441 TCCTTGTTCTGCGACGGTTCCATTCTTTTTGGTTTTTAGTTTGCAAATCTTGATCGTGGA 1500
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1441 TCCTTGTTCTGCGACGGTTCCATTCTTTTTGGTTTTTAGTTTGCAAATCTTGATCGTGGA 1500

Qy       1501 GTTGAGAAAAAGGGCGGTTCGTTGTCTTGAGGTGTTCTTGTTGATTGTTCGTCATGGAAA 1560
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1501 GTTGAGAAAAAGGGCGGTTCGTTGTCTTGAGGTGTTCTTGTTGATTGTTCGTCATGGAAA 1560

Qy       1561 ATAATGATGCACTTGACATTGGAGCCGTGTCGTCCCCATATCCTTCGCAATCTGAAGGAG 1620
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1561 ATAATGATGCACTTGACATTGGAGCCGTGTCGTCCCCATATCCTTCGCAATCTGAAGGAG 1620

Qy       1621 TGTCTACGCCATTGCCGCAAGTAACATCACCGAGCTTCGACAATGCAGCCTCACCCGTGG 1680
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1621 TGTCTACGCCATTGCCGCAAGTAACATCACCGAGCTTCGACAATGCAGCCTCACCCGTGG 1680

Qy       1681 CCGGGCGGAGGGCCGTACGGCAGACCCCTACATCTGCAGTTCGAAGGAGAGGGAGAGAAA 1740
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1681 CCGGGCGGAGGGCCGTACGGCAGACCCCTACATCTGCAGTTCGAAGGAGAGGGAGAGAAA 1740

Qy       1741 CGGATTCCGCTCGTCGTAGGAGGAGTCGATCTCGCAGTTTAGGCAATTCTGTTTATAGTT 1800
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1741 CGGATTCCGCTCGTCGTAGGAGGAGTCGATCTCGCAGTTTAGGCAATTCTGTTTATAGTT 1800

Qy       1801 CCCCTTACGATGCGGGGACTCCTGGAACTCCTGGAACTCCAGTGGCTACTCCGGTTTACG 1860
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1801 CCCCTTACGATGCGGGGACTCCTGGAACTCCTGGAACTCCAGTGGCTACTCCGGTTTACG 1860

Qy       1861 CTACCCCAGTCGGTACACCTATGGGTACCCCATCGTTCCATCGTGGCACGCCACAGTACA 1920
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1861 CTACCCCAGTCGGTACACCTATGGGTACCCCATCGTTCCATCGTGGCACGCCACAGTACA 1920

Qy       1921 AACAGCGCAGTGAGCTTGGTTCCCAGGGGAAGCCTCTACATCGGAGACGTCGATCTCAAT 1980
```

```
                ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db         1921 AACAGCGCAGTGAGCTTGGTTCCCAGGGGAAGCCTCTACATCGGAGACGTCGATCTCAAT 1980

Qy         1981 CCAGAGAACCCGGGCATCGATCTCCTTCAAGGGAACCTAGTGCTGATGGGCGTCCCTCTG 2040
                ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db         1981 CCAGAGAACCCGGGCATCGATCTCCTTCAAGGGAACCTAGTGCTGATGGGCGTCCCTCTG 2040

Qy         2041 AATCTGCTGAGCCAGATGACACTTTGGGTGGAGAATATGCTTATGTTTGGGGGACGAATG 2100
                ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db         2041 AATCTGCTGAGCCAGATGACACTTTGGGTGGAGAATATGCTTATGTTTGGGGGACGAATG 2100

Qy         2101 TTAACATTCCAGATGTGCTTAGGGCGATTCGTCGATTTCTCCACAATTATCGTTCGAGTG 2160
                ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db         2101 TTAACATTCCAGATGTGCTTAGGGCGATTCGTCGATTTCTCCACAATTATCGTTCGAGTG 2160

Qy         2161 CTCATGATCTTAATTCCAAGTACATCCAGATCATAGAGGAGACTGTGGAGCGTGAGGAGG 2220
                ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db         2161 CTCATGATCTTAATTCCAAGTACATCCAGATCATAGAGGAGACTGTGGAGCGTGAGGAGG 2220

Qy         2221 ATACTCTAAATATCGACATGTCAGACATTTATGACCATGATCCTGATCTATACGCAAAAA 2280
                ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db         2221 ATACTCTAAATATCGACATGTCAGACATTTATGACCATGATCCTGATCTATACGCAAAAA 2280

Qy         2281 TTGTTCGATACCCACTCGACATCATCCCCCTGTTGGACACTGAGTGTCAGGAAGTTGCTA 2340
                ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db         2281 TTGTTCGATACCCACTCGACATCATCCCCCTGTTGGACACTGAGTGTCAGGAAGTTGCTA 2340

Qy         2341 CCTCTTTACTACCAACGTTTGAGAAGCATATTGAGGCCAGACCTTTCAATCTCAAAGCAT 2400
                ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db         2341 CCTCTTTACTACCAACGTTTGAGAAGCATATTGAGGCCAGACCTTTCAATCTCAAAGCAT 2400

Qy         2401 CGGTGCACATGCGTGAACTCAACCCTTCAGATATAGACAAATTGGTTTCTGTTAAAGGAA 2460
                ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db         2401 CGGTGCACATGCGTGAACTCAACCCTTCAGATATAGACAAATTGGTTTCTGTTAAAGGAA 2460

Qy         2461 TGGTTATCCGGTGCAGTTCTATCATACCTGAAATTAAGGGGGCCTTCTTCAAATGTTTAG 2520
                ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db         2461 TGGTTATCCGGTGCAGTTCTATCATACCTGAAATTAAGGGGGCCTTCTTCAAATGTTTAG 2520

Qy         2521 TGTGTGGTCACTCGCCTCCGCTAGTTACAGTTGTTAAAGGGCGGGTTGAGGAGCCAACAA 2580
                ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db         2521 TGTGTGGTCACTCGCCTCCGCTAGTTACAGTTGTTAAAGGGCGGGTTGAGGAGCCAACAA 2580

Qy         2581 GGTGTGAAAAGCCAGAATGTGCAGCACGGAATGCTATGTCTCTTATTCACAATCGATGCA 2640
                ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db         2581 GGTGTGAAAAGCCAGAATGTGCAGCACGGAATGCTATGTCTCTTATTCACAATCGATGCA 2640

Qy         2641 CTTTTGCAAATAAGCAGATAGTGCGTCTTCAAGAAACTCCAGATGCCATTCCTGAAGGAG 2700
                ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db         2641 CTTTTGCAAATAAGCAGATAGTGCGTCTTCAAGAAACTCCAGATGCCATTCCTGAAGGAG 2700

Qy         2701 AGACTCCACACACAGTCAGCATGTGTTTATACAACACTATGGTTGATGCTGTGAAGCCTG 2760
                ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db         2701 AGACTCCACACACAGTCAGCATGTGTTTATACAACACTATGGTTGATGCTGTGAAGCCTG 2760

Qy         2761 GAGATCGTATTGAGGTAACAGGAGTTTTCAAGGCCATGGCAGTTCGAGTTGGTCCGAATC 2820
                ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db         2761 GAGATCGTATTGAGGTAACAGGAGTTTTCAAGGCCATGGCAGTTCGAGTTGGTCCGAATC 2820

Qy         2821 AACGAACATTACGAGCATTGTATAAGACCTACATCGATTGCGTGCACGTCAAGAAGTCTG 2880
                ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db         2821 AACGAACATTACGAGCATTGTATAAGACCTACATCGATTGCGTGCACGTCAAGAAGTCTG 2880

Qy         2881 ACAGGGGTCGACTGCAAACTGAAGATCCTATGGAGATGGATAAGGAGAATGATATGTATG 2940
                ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db         2881 ACAGGGGTCGACTGCAAACTGAAGATCCTATGGAGATGGATAAGGAGAATGATATGTATG 2940

Qy         2941 CTGGGTATCATGAAAGTGATACTTCAGAAGCTGCTAATGAAGCAAAGATTCAAAAACTTA 3000
                ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db         2941 CTGGGTATCATGAAAGTGATACTTCAGAAGCTGCTAATGAAGCAAAGATTCAAAAACTTA 3000

Qy         3001 AAGAGCTGTCCAAGCTCCCGGGCATTTATGATAGACTTTCAAGGTCGCTGGCTCCAAGCA 3060
                ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db         3001 AAGAGCTGTCCAAGCTCCCGGGCATTTATGATAGACTTTCAAGGTCGCTGGCTCCAAGCA 3060

Qy         3061 TTTGGGAGCTTGAAGATATTAAAAAGGGTCTTCTTTGCCAGCTCTTTGGTGGAAGGCTA 3120
                ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db         3061 TTTGGGAGCTTGAAGATATTAAAAAGGGTCTTCTTTGCCAGCTCTTTGGTGGAAGGCTA 3120

Qy         3121 AGAAAATTCCATCTGGAGCATCTTTCCGAGGTGACATCAATGTTTTACTTGTTGGGGACC 3180
                ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db         3121 AGAAAATTCCATCTGGAGCATCTTTCCGAGGTGACATCAATGTTTTACTTGTTGGGGACC 3180

Qy         3181 CTGGTACCAGTAAATCTCAGCTGCTTCAGTATGTGCACAAGATAGCTCCTCGTGGAATCT 3240
                ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db         3181 CTGGTACCAGTAAATCTCAGCTGCTTCAGTATGTGCACAAGATAGCTCCTCGTGGAATCT 3240

Qy         3241 ACACTAGTGGGCGAGGAAGTTCGGCGGTTGGGCTGACAGCGTATGTAACGAAGGATCCAG 3300
                ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db         3241 ACACTAGTGGGCGAGGAAGTTCGGCGGTTGGGCTGACAGCGTATGTAACGAAGGATCCAG 3300

Qy         3301 AAACTCGAGAGACGGTATTGGAGAGCGGAGCTTTGGTTCTTAGTGATCGTGGGATATGCT 3360
                ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db         3301 AAACTCGAGAGACGGTATTGGAGAGCGGAGCTTTGGTTCTTAGTGATCGTGGGATATGCT 3360

Qy         3361 GTATCGATGAGTTCGACAAAATGTCTGATAATGCCCGAAGCATGCTTCATGAGGTAATGG 3420
                ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db         3361 GTATCGATGAGTTCGACAAAATGTCTGATAATGCCCGAAGCATGCTTCATGAGGTAATGG 3420

Qy         3421 AGCAACAAACGGTATCTGTAGCCAAAGGGGGTATCATTGCCTCGCTGAACGCTCGGACGT 3480
                ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db         3421 AGCAACAAACGGTATCTGTAGCCAAAGGGGGTATCATTGCCTCGCTGAACGCTCGGACGT 3480

Qy         3481 CTGTCCTTGCATGTGCAAATCCTAGTGGGTCCCGATACAATGCGCGCCTTTCTGTGATTG 3540
                ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db         3481 CTGTCCTTGCATGTGCAAATCCTAGTGGGTCCCGATACAATGCGCGCCTTTCTGTGATTG 3540

Qy         3541 ATAACATCCAGCTTCCTCCAACTCTACTTTCTAGATTTGATTTAATTTACTTAATGCTCG 3600
                ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db         3541 ATAACATCCAGCTTCCTCCAACTCTACTTTCTAGATTTGATTTAATTTACTTAATGCTCG 3600

Qy         3601 ACAAACCAGACGAGCAAAACGATCGTCGTCTCGCCAGGCATCTCGTGGCTTTACACTATG 3660
                ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db         3601 ACAAACCAGACGAGCAAAACGATCGTCGTCTCGCCAGGCATCTCGTGGCTTTACACTATG 3660

Qy         3661 AAAACTATGAAGTTTCAAAGCAGGACGCCTTAGATCTACAAACACTTACCGCGTATATCA 3720
                ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db         3661 AAAACTATGAAGTTTCAAAGCAGGACGCCTTAGATCTACAAACACTTACCGCGTATATCA 3720

Qy         3721 CCTATGCTCGTCAGCATGTACATCCTACATTAAGTGATGAAGCTGCTGAAGATTTGATTA 3780
                ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db         3721 CCTATGCTCGTCAGCATGTACATCCTACATTAAGTGATGAAGCTGCTGAAGATTTGATTA 3780

Qy         3781 ATGGCTATGTTGAGATGCGCCAAAAGGGCAACTTTCCTGGAAGCAGTAAAAAGGTGATAA 3840
                ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db         3781 ATGGCTATGTTGAGATGCGCCAAAAGGGCAACTTTCCTGGAAGCAGTAAAAAGGTGATAA 3840

Qy         3841 CAGCCACACCTCGGCAACTCGAAAGTATGATTCGTATCAGTGAAGCCCTAGCTCGAATGA 3900
                ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db         3841 CAGCCACACCTCGGCAACTCGAAAGTATGATTCGTATCAGTGAAGCCCTAGCTCGAATGA 3900

Qy         3901 GATTTTCTGAAGTGGTAGAGAAAGTTGATGCAGCAGAAGCTGTGCGCCTTTTAGACGTCG 3960
                ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db         3901 GATTTTCTGAAGTGGTAGAGAAAGTTGATGCAGCAGAAGCTGTGCGCCTTTTAGACGTCG 3960

Qy         3961 CTTTGCAGCAATCTGCTACTGATCATGCAACAGGTACGATAGACATGGATCTTATCACGA 4020
                ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db         3961 CTTTGCAGCAATCTGCTACTGATCATGCAACAGGTACGATAGACATGGATCTTATCACGA 4020

Qy         4021 CTGGAGTGTCGGCCAGCGAGCGTATTCGTCGGGCCAACTTGCTAGCTGCTCTGCGAGAGC 4080
                ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db         4021 CTGGAGTGTCGGCCAGCGAGCGTATTCGTCGGGCCAACTTGCTAGCTGCTCTGCGAGAGC 4080

Qy         4081 TTATAGCAGATAAAATTTCACCTGGCAGCTCCTCTGGCTTGAAGACCAGTCAGCTTCTTG 4140
                ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db         4081 TTATAGCAGATAAAATTTCACCTGGCAGCTCCTCTGGCTTGAAGACCAGTCAGCTTCTTG 4140
```

```
Qy       4141 AGGATATCCGGAGCCAAAGCAGTGTGGACGTTAGTTTGCAGGATATTAAAAATGCTCTGG 4200
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       4141 AGGATATCCGGAGCCAAAGCAGTGTGGACGTTAGTTTGCAGGATATTAAAAATGCTCTGG 4200

Qy       4201 GTAGCCTCCAAGGAGAAGGCTTTCTTACTGTCCATGGTGACATAGTCAAGAGAGTTTGAG 4260
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       4201 GTAGCCTCCAAGGAGAAGGCTTTCTTACTGTCCATGGTGACATAGTCAAGAGAGTTTGAG 4260

Qy       4261 ACAGTTTCTAACTGTTCGAATCCATGAGCTATAACTCTGAACGAAAGGGAAAACCTCCAG 4320
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       4261 ACAGTTTCTAACTGTTCGAATCCATGAGCTATAACTCTGAACGAAAGGGAAAACCTCCAG 4320

Qy       4321 TTTCCCATGCGCAATTCCCAGAGCTCGC 4348
              ||||||||||||||||||||||||||||
Db       4321 TTTCCCATGCGCAATTCCCAGAGCTCGC 4348


RESULT 2
AAD60813
ID   AAD60813 standard; cDNA; 4348 BP.
XX
AC   AAD60813;
XX
DT   15-JAN-2004  (first entry)
XX
DE   Physcomitrella patens cell cycle protein (CC-2) cDNA.
XX
KW   Cell cycle stress-related protein; CCSRP; metal; cell cycle protein; CC;
KW   transgenic plant; environmental stress; stress tolerance; salinity;
KW   drought; temperature; chemical; oxidative stress; gene; ss.
XX
OS   Physcomitrella patens.
XX
FH   Key             Location/Qualifiers
FT   CDS             147. .4259
FT                   /*tag=  a
FT                   /product= "CC-2 protein"
FT                   /transl_except= (pos:1404. .2816, aa:Gly-Pro)
XX
PN   US2003097675-A1.
XX
PD   22-MAY-2003.
XX
PF   06-APR-2001; 2001US-00828062.
XX
PR   07-APR-2000; 2000US-0196001P.
XX
PA   (SILV/) COSTA E SILVA O.
PA   (BOHN/) BOHNERT H J.
PA   (THIE/) THIELEN N V.
PA   (CHEN/) CHEN R.
PA   (SARR/) SARRIA-MILLAN R.
XX
PI   Costa E SilvaO,  Bohnert HJ,  Thielen NV,  Chen R,  Sarria-Millan R;
XX
DR   WPI; 2003-765533/72.
DR   P-PSDB; AAE39980.
XX
PT   Novel cell cycle stress-related protein capable of conferring stress
PT   tolerance such as tolerance towards salinity, drought, temperature,
PT   chemical, pathogens, to plants upon over-expression.
XX
PS   Claim 16; Fig 2B; 62pp; English.
XX
CC   The invention relates to an isolated cell cycle stress-related protein
CC   (CCSRP) which is chosen from a cell cycle-1 protein (CC-1), CC-2, CC-3
CC   and its orthologues. Recombinant expression vector is useful for
CC   producing a transgenic plant containing a CCSRP coding nucleic acid,
CC   where expression of the nucleic acid in the plant results in increased
CC   tolerance to environmental stress as compared to a wild type variety of
CC   the plant which involves transforming a plant cell with the recombinant
CC   expression vector, generating from the plant cell a transgenic plant with
CC   an increased tolerance to environmental stress as compared to a wild type
CC   variety of the plant. CCSRP is useful for conferring stress tolerance
CC   such as tolerance towards salinity, drought, temperature, metal,
CC   chemical, pathogens and oxidative stress or their combinations to plants.
CC   The present sequence is Physcomitrella patens CC-2 cDNA
XX
SQ   Sequence 4348 BP; 1120 A; 987 C; 1093 G; 1148 T; 0 U; 0 Other;

  Query Match             100.0%;  Score 4348;  DB 10;  Length 4348;
  Best Local Similarity   100.0%;  Pred. No. 0;
  Matches 4348;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy          1 ATGGCGCGCCGCACTCACGTGAGGAATTGCACCTCCTTGTTCTGCGACGGTTCCATTCTT 60
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db          1 ATGGCGCGCCGCACTCACGTGAGGAATTGCACCTCCTTGTTCTGCGACGGTTCCATTCTT 60

Qy         61 TTTGGTTTTTAGTTTGCAAATCTTGATCGTGGAGTTGAGAAAAAGGGCGGTTCGTTGTCT 120
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db         61 TTTGGTTTTTAGTTTGCAAATCTTGATCGTGGAGTTGAGAAAAAGGGCGGTTCGTTGTCT 120

Qy        121 TGAGGTGTTCTTGTTGATTGTTCGTCATGGAAAATAATGATGCACTTGACATTGGAGCCG 180
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        121 TGAGGTGTTCTTGTTGATTGTTCGTCATGGAAAATAATGATGCACTTGACATTGGAGCCG 180

Qy        181 TGTCGTCCCCATATCCTTCGCAATCTGAAGGAGTGTCTACGCCATTGCCGCAAGTAACAT 240
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        181 TGTCGTCCCCATATCCTTCGCAATCTGAAGGAGTGTCTACGCCATTGCCGCAAGTAACAT 240

Qy        241 CACCGAGCTTCGACAATGCAGCCTCACCCGTGGCCGGGCGGAGGGCCGTACGGCAGACCC 300
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        241 CACCGAGCTTCGACAATGCAGCCTCACCCGTGGCCGGGCGGAGGGCCGTACGGCAGACCC 300

Qy        301 CTACATCTGCAGTTCGAAGGAGAGGGAGAGAAACGGATTCCGCTCGTCGTAGGAGGAGTC 360
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        301 CTACATCTGCAGTTCGAAGGAGAGGGAGAGAAACGGATTCCGCTCGTCGTAGGAGGAGTC 360

Qy        361 GATCTCGCAGTTTAGGCAATTCTGTTTATAGTTCCCCTTACGATGCGGGGACTCCTGGAA 420
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        361 GATCTCGCAGTTTAGGCAATTCTGTTTATAGTTCCCCTTACGATGCGGGGACTCCTGGAA 420

Qy        421 CTCCTGGAACTCCAGTGGCTACTCCGGTTTACGCTACCCCAGTCGGTACACCTATGGGTA 480
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        421 CTCCTGGAACTCCAGTGGCTACTCCGGTTTACGCTACCCCAGTCGGTACACCTATGGGTA 480

Qy        481 CCCCATCGTTCCATCGTGGCACGCCACAGTACAAACAGCGCAGTGAGCTTGGTTCCCAGG 540
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        481 CCCCATCGTTCCATCGTGGCACGCCACAGTACAAACAGCGCAGTGAGCTTGGTTCCCAGG 540

Qy        541 GGAAGCCTCTACATCGGAGACGTCGATCTCAATCCAGAGAACCCGGGCATCGATCTCCTT 600
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        541 GGAAGCCTCTACATCGGAGACGTCGATCTCAATCCAGAGAACCCGGGCATCGATCTCCTT 600

Qy        601 CAAGGGAACCTAGTGCTGATGGGCGTCCCTCTGAATCTGCTGAGCCAGATGACACTTTGG 660
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        601 CAAGGGAACCTAGTGCTGATGGGCGTCCCTCTGAATCTGCTGAGCCAGATGACACTTTGG 660

Qy        661 GTGGAGAATATGCTTATGTTTGGGGGACGAATGTTAACATTCCAGATGTGCTTAGGGCGA 720
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        661 GTGGAGAATATGCTTATGTTTGGGGGACGAATGTTAACATTCCAGATGTGCTTAGGGCGA 720

Qy        721 TTCGTCGATTTCTCCACAATTATCGTTCGAGTGCTCATGATCTTAATTCCAAGTACATCC 780
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        721 TTCGTCGATTTCTCCACAATTATCGTTCGAGTGCTCATGATCTTAATTCCAAGTACATCC 780

Qy        781 AGATCATAGAGGAGACTGTGGAGCGTGAGGAGGATACTCTAAATATCGACATGTCAGACA 840
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        781 AGATCATAGAGGAGACTGTGGAGCGTGAGGAGGATACTCTAAATATCGACATGTCAGACA 840

Qy        841 TTTATGACCATGATCCTGATCTATACGCAAAAATTGTTCGATACCCACTCGACATCATCC 900
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        841 TTTATGACCATGATCCTGATCTATACGCAAAAATTGTTCGATACCCACTCGACATCATCC 900

Qy        901 CCCTGTTGGACACTGAGTGTCAGGAAGTTGCTACCTCTTTACTACCAACGTTTGAGAAGC 960
```

```
               ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        901 CCCTGTTGGACACTGAGTGTCAGGAAGTTGCTACCTCTTTACTACCAACGTTTGAGAAGC 960

Qy        961 ATATTGAGGCCAGACCTTTCAATCTCAAAGCATCGGTGCACATGCGTGAACTCAACCCTT 1020
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        961 ATATTGAGGCCAGACCTTTCAATCTCAAAGCATCGGTGCACATGCGTGAACTCAACCCTT 1020

Qy       1021 CAGATATAGACAAATTGGTTTCTGTTAAAGGAATGGTTATCCGGTGCAGTTCTATCATAC 1080
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1021 CAGATATAGACAAATTGGTTTCTGTTAAAGGAATGGTTATCCGGTGCAGTTCTATCATAC 1080

Qy       1081 CTGAAATTAAGGGGGCCTTCTTCAAATGTTTAGTGTGTGGTCACTCGCCTCCGCTAGTTA 1140
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1081 CTGAAATTAAGGGGGCCTTCTTCAAATGTTTAGTGTGTGGTCACTCGCCTCCGCTAGTTA 1140

Qy       1141 CAGTTGTTAAAGGGCGGGTTGAGGAGCCAACAAGGTGTGAAAAGCCAGAATGTGCAGCAC 1200
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1141 CAGTTGTTAAAGGGCGGGTTGAGGAGCCAACAAGGTGTGAAAAGCCAGAATGTGCAGCAC 1200

Qy       1201 GGAATGCTATGTCTCTTATTCACAATCGATGCACTTTTGCAAATAAGCAGATAGTGCGTC 1260
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1201 GGAATGCTATGTCTCTTATTCACAATCGATGCACTTTTGCAAATAAGCAGATAGTGCGTC 1260

Qy       1261 TTCAAGAAACTCCAGATGCCATTCCTGAAGGAGAGACTCCACACACAGTCAGCATGTGTT 1320
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1261 TTCAAGAAACTCCAGATGCCATTCCTGAAGGAGAGACTCCACACACAGTCAGCATGTGTT 1320

Qy       1321 TATACAACACTATGGTTGATGCTGTGAAGCCTGGAGATCGTATTGAGGTAACAGGAGTTT 1380
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1321 TATACAACACTATGGTTGATGCTGTGAAGCCTGGAGATCGTATTGAGGTAACAGGAGTTT 1380

Qy       1381 TCAAGGCCATGGCAGTTCGAGTTGGTCATGGCGCGCCGCACTCACGTGAGGAATTGCACC 1440
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1381 TCAAGGCCATGGCAGTTCGAGTTGGTCATGGCGCGCCGCACTCACGTGAGGAATTGCACC 1440

Qy       1441 TCCTTGTTCTGCGACGGTTCCATTCTTTTTGGTTTTTAGTTTGCAAATCTTGATCGTGGA 1500
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1441 TCCTTGTTCTGCGACGGTTCCATTCTTTTTGGTTTTTAGTTTGCAAATCTTGATCGTGGA 1500

Qy       1501 GTTGAGAAAAAGGGCGGTTCGTTGTCTTGAGGTGTTCTTGTTGATTGTTCGTCATGGAAA 1560
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1501 GTTGAGAAAAAGGGCGGTTCGTTGTCTTGAGGTGTTCTTGTTGATTGTTCGTCATGGAAA 1560

Qy       1561 ATAATGATGCACTTGACATTGGAGCCGTGTCGTCCCCATATCCTTCGCAATCTGAAGGAG 1620
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1561 ATAATGATGCACTTGACATTGGAGCCGTGTCGTCCCCATATCCTTCGCAATCTGAAGGAG 1620

Qy       1621 TGTCTACGCCATTGCCGCAAGTAACATCACCGAGCTTCGACAATGCAGCCTCACCCGTGG 1680
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1621 TGTCTACGCCATTGCCGCAAGTAACATCACCGAGCTTCGACAATGCAGCCTCACCCGTGG 1680

Qy       1681 CCGGGCGGAGGGCCGTACGGCAGACCCCTACATCTGCAGTTCGAAGGAGAGGGAGAGAAA 1740
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1681 CCGGGCGGAGGGCCGTACGGCAGACCCCTACATCTGCAGTTCGAAGGAGAGGGAGAGAAA 1740

Qy       1741 CGGATTCCGCTCGTCGTAGGAGGAGTCGATCTCGCAGTTTAGGCAATTCTGTTTATAGTT 1800
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1741 CGGATTCCGCTCGTCGTAGGAGGAGTCGATCTCGCAGTTTAGGCAATTCTGTTTATAGTT 1800

Qy       1801 CCCCTTACGATGCGGGGACTCCTGGAACTCCTGGAACTCCAGTGGCTACTCCGGTTTACG 1860
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1801 CCCCTTACGATGCGGGGACTCCTGGAACTCCTGGAACTCCAGTGGCTACTCCGGTTTACG 1860

Qy       1861 CTACCCCAGTCGGTACACCTATGGGTACCCCATCGTTCCATCGTGGCACGCCACAGTACA 1920
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1861 CTACCCCAGTCGGTACACCTATGGGTACCCCATCGTTCCATCGTGGCACGCCACAGTACA 1920

Qy       1921 AACAGCGCAGTGAGCTTGGTTCCCAGGGGAAGCCTCTACATCGGAGACGTCGATCTCAAT 1980
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1921 AACAGCGCAGTGAGCTTGGTTCCCAGGGGAAGCCTCTACATCGGAGACGTCGATCTCAAT 1980

Qy       1981 CCAGAGAACCCGGGCATCGATCTCCTTCAAGGGAACCTAGTGCTGATGGGCGTCCCTCTG 2040
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1981 CCAGAGAACCCGGGCATCGATCTCCTTCAAGGGAACCTAGTGCTGATGGGCGTCCCTCTG 2040

Qy       2041 AATCTGCTGAGCCAGATGACACTTTGGGTGGAGAATATGCTTATGTTTGGGGGACGAATG 2100
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       2041 AATCTGCTGAGCCAGATGACACTTTGGGTGGAGAATATGCTTATGTTTGGGGGACGAATG 2100

Qy       2101 TTAACATTCCAGATGTGCTTAGGGCGATTCGTCGATTTCTCCACAATTATCGTTCGAGTG 2160
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       2101 TTAACATTCCAGATGTGCTTAGGGCGATTCGTCGATTTCTCCACAATTATCGTTCGAGTG 2160

Qy       2161 CTCATGATCTTAATTCCAAGTACATCCAGATCATAGAGGAGACTGTGGAGCGTGAGGAGG 2220
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       2161 CTCATGATCTTAATTCCAAGTACATCCAGATCATAGAGGAGACTGTGGAGCGTGAGGAGG 2220

Qy       2221 ATACTCTAAATATCGACATGTCAGACATTTATGACCATGATCCTGATCTATACGCAAAAA 2280
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       2221 ATACTCTAAATATCGACATGTCAGACATTTATGACCATGATCCTGATCTATACGCAAAAA 2280

Qy       2281 TTGTTCGATACCCACTCGACATCATCCCCCTGTTGGACACTGAGTGTCAGGAAGTTGCTA 2340
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       2281 TTGTTCGATACCCACTCGACATCATCCCCCTGTTGGACACTGAGTGTCAGGAAGTTGCTA 2340

Qy       2341 CCTCTTTACTACCAACGTTTGAGAAGCATATTGAGGCCAGACCTTTCAATCTCAAAGCAT 2400
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       2341 CCTCTTTACTACCAACGTTTGAGAAGCATATTGAGGCCAGACCTTTCAATCTCAAAGCAT 2400

Qy       2401 CGGTGCACATGCGTGAACTCAACCCTTCAGATATAGACAAATTGGTTTCTGTTAAAGGAA 2460
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       2401 CGGTGCACATGCGTGAACTCAACCCTTCAGATATAGACAAATTGGTTTCTGTTAAAGGAA 2460

Qy       2461 TGGTTATCCGGTGCAGTTCTATCATACCTGAAATTAAGGGGGCCTTCTTCAAATGTTTAG 2520
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       2461 TGGTTATCCGGTGCAGTTCTATCATACCTGAAATTAAGGGGGCCTTCTTCAAATGTTTAG 2520

Qy       2521 TGTGTGGTCACTCGCCTCCGCTAGTTACAGTTGTTAAAGGGCGGGTTGAGGAGCCAACAA 2580
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       2521 TGTGTGGTCACTCGCCTCCGCTAGTTACAGTTGTTAAAGGGCGGGTTGAGGAGCCAACAA 2580

Qy       2581 GGTGTGAAAAGCCAGAATGTGCAGCACGGAATGCTATGTCTCTTATTCACAATCGATGCA 2640
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       2581 GGTGTGAAAAGCCAGAATGTGCAGCACGGAATGCTATGTCTCTTATTCACAATCGATGCA 2640

Qy       2641 CTTTTGCAAATAAGCAGATAGTGCGTCTTCAAGAAACTCCAGATGCCATTCCTGAAGGAG 2700
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       2641 CTTTTGCAAATAAGCAGATAGTGCGTCTTCAAGAAACTCCAGATGCCATTCCTGAAGGAG 2700

Qy       2701 AGACTCCACACACAGTCAGCATGTGTTTATACAACACTATGGTTGATGCTGTGAAGCCTG 2760
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       2701 AGACTCCACACACAGTCAGCATGTGTTTATACAACACTATGGTTGATGCTGTGAAGCCTG 2760

Qy       2761 GAGATCGTATTGAGGTAACAGGAGTTTTCAAGGCCATGGCAGTTCGAGTTGGTCCGAATC 2820
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       2761 GAGATCGTATTGAGGTAACAGGAGTTTTCAAGGCCATGGCAGTTCGAGTTGGTCCGAATC 2820

Qy       2821 AACGAACATTACGAGCATTGTATAAGACCTACATCGATTGCGTGCACGTCAAGAAGTCTG 2880
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       2821 AACGAACATTACGAGCATTGTATAAGACCTACATCGATTGCGTGCACGTCAAGAAGTCTG 2880

Qy       2881 ACAGGGGTCGACTGCAAACTGAAGATCCTATGGAGATGGATAAGGAGAATGATATGTATG 2940
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       2881 ACAGGGGTCGACTGCAAACTGAAGATCCTATGGAGATGGATAAGGAGAATGATATGTATG 2940

Qy       2941 CTGGGTATCATGAAAGTGATACTTCAGAAGCTGCTAATGAAGCAAAGATTCAAAAACTTA 3000
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       2941 CTGGGTATCATGAAAGTGATACTTCAGAAGCTGCTAATGAAGCAAAGATTCAAAAACTTA 3000

Qy       3001 AAGAGCTGTCCAAGCTCCCGGGCATTTATGATAGACTTTCAAGGTCGCTGGCTCCAAGCA 3060
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       3001 AAGAGCTGTCCAAGCTCCCGGGCATTTATGATAGACTTTCAAGGTCGCTGGCTCCAAGCA 3060

Qy       3061 TTTGGGAGCTTGAAGATATTAAAAAGGGTCTTCTTTGCCAGCTCTTTGGTGGGAAGGCTA 3120
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       3061 TTTGGGAGCTTGAAGATATTAAAAAGGGTCTTCTTTGCCAGCTCTTTGGTGGGAAGGCTA 3120
```

```
Qy      3121 AGAAAATTCCATCTGGAGCATCTTTCCGAGGTGACATCAATGTTTTACTTGTTGGGGACC 3180
             ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      3121 AGAAAATTCCATCTGGAGCATCTTTCCGAGGTGACATCAATGTTTTACTTGTTGGGGACC 3180

Qy      3181 CTGGTACCAGTAAATCTCAGCTGCTTCAGTATGTGCACAAGATAGCTCCTCGTGGAATCT 3240
             ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      3181 CTGGTACCAGTAAATCTCAGCTGCTTCAGTATGTGCACAAGATAGCTCCTCGTGGAATCT 3240

Qy      3241 ACACTAGTGGGCGAGGAAGTTCGGCGGTTGGGCTGACAGCGTATGTAACGAAGGATCCAG 3300
             ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      3241 ACACTAGTGGGCGAGGAAGTTCGGCGGTTGGGCTGACAGCGTATGTAACGAAGGATCCAG 3300

Qy      3301 AAACTCGAGAGACGGTATTGGAGAGCGGAGCTTTGGTTCTTAGTGATCGTGGGATATGCT 3360
             ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      3301 AAACTCGAGAGACGGTATTGGAGAGCGGAGCTTTGGTTCTTAGTGATCGTGGGATATGCT 3360

Qy      3361 GTATCGATGAGTTCGACAAAATGTCTGATAATGCCCGAAGCATGCTTCATGAGGTAATGG 3420
             ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      3361 GTATCGATGAGTTCGACAAAATGTCTGATAATGCCCGAAGCATGCTTCATGAGGTAATGG 3420

Qy      3421 AGCAACAAACGGTATCTGTAGCCAAAGGGGGTATCATTGCCTCGCTGAACGCTCGGACGT 3480
             ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      3421 AGCAACAAACGGTATCTGTAGCCAAAGGGGGTATCATTGCCTCGCTGAACGCTCGGACGT 3480

Qy      3481 CTGTCCTTGCATGTGCAAATCCTAGTGGGTCCCGATACAATGCGCGCCTTTCTGTGATTG 3540
             ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      3481 CTGTCCTTGCATGTGCAAATCCTAGTGGGTCCCGATACAATGCGCGCCTTTCTGTGATTG 3540

Qy      3541 ATAACATCCAGCTTCCTCCAACTCTACTTTCTAGATTTGATTTAATTTACTTAATGCTCG 3600
             ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      3541 ATAACATCCAGCTTCCTCCAACTCTACTTTCTAGATTTGATTTAATTTACTTAATGCTCG 3600

Qy      3601 ACAAACCAGACGAGCAAAACGATCGTCGTCTCGCCAGGCATCTCGTGGCTTTACACTATG 3660
             ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      3601 ACAAACCAGACGAGCAAAACGATCGTCGTCTCGCCAGGCATCTCGTGGCTTTACACTATG 3660

Qy      3661 AAAACTATGAAGTTTCAAAGCAGGACGCCTTAGATCTACAAACACTTACCGCGTATATCA 3720
             ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      3661 AAAACTATGAAGTTTCAAAGCAGGACGCCTTAGATCTACAAACACTTACCGCGTATATCA 3720

Qy      3721 CCTATGCTCGTCAGCATGTACATCCTACATTAAGTGATGAAGCTGCTGAAGATTTGATTA 3780
             ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      3721 CCTATGCTCGTCAGCATGTACATCCTACATTAAGTGATGAAGCTGCTGAAGATTTGATTA 3780

Qy      3781 ATGGCTATGTTGAGATGCGCCAAAAGGGCAACTTTCCTGGAAGCAGTAAAAAGGTGATAA 3840
             ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      3781 ATGGCTATGTTGAGATGCGCCAAAAGGGCAACTTTCCTGGAAGCAGTAAAAAGGTGATAA 3840

Qy      3841 CAGCCACACCTCGGCAACTCGAAAGTATGATTCGTATCAGTGAAGCCCTAGCTCGAATGA 3900
             ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      3841 CAGCCACACCTCGGCAACTCGAAAGTATGATTCGTATCAGTGAAGCCCTAGCTCGAATGA 3900

Qy      3901 GATTTTCTGAAGTGGTAGAGAAAGTTGATGCAGCAGAAGCTGTGCGCCTTTTAGACGTCG 3960
             ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      3901 GATTTTCTGAAGTGGTAGAGAAAGTTGATGCAGCAGAAGCTGTGCGCCTTTTAGACGTCG 3960

Qy      3961 CTTTGCAGCAATCTGCTACTGATCATGCAACAGGTACGATAGACATGGATCTTATCACGA 4020
             ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      3961 CTTTGCAGCAATCTGCTACTGATCATGCAACAGGTACGATAGACATGGATCTTATCACGA 4020

Qy      4021 CTGGAGTGTCGGCCAGCGAGCGTATTCGTCGGGCCAACTTGCTAGCTGCTCTGCGAGAGC 4080
             ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      4021 CTGGAGTGTCGGCCAGCGAGCGTATTCGTCGGGCCAACTTGCTAGCTGCTCTGCGAGAGC 4080

Qy      4081 TTATAGCAGATAAAATTTCACCTGGCAGCTCCTCTGGCTTGAAGACCAGTCAGCTTCTTG 4140
             ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      4081 TTATAGCAGATAAAATTTCACCTGGCAGCTCCTCTGGCTTGAAGACCAGTCAGCTTCTTG 4140

Qy      4141 AGGATATCCGGAGCCAAAGCAGTGTGGACGTTAGTTTGCAGGATATTAAAAATGCTCTGG 4200
             ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      4141 AGGATATCCGGAGCCAAAGCAGTGTGGACGTTAGTTTGCAGGATATTAAAAATGCTCTGG 4200

Qy      4201 GTAGCCTCCAAGGAGAAGGCTTTCTTACTGTCCATGGTGACATAGTCAAGAGAGTTTGAG 4260
             ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      4201 GTAGCCTCCAAGGAGAAGGCTTTCTTACTGTCCATGGTGACATAGTCAAGAGAGTTTGAG 4260

Qy      4261 ACAGTTTCTAACTGTTCGAATCCATGAGCTATAACTCTGAACGAAAGGGAAAACCTCCAG 4320
             ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      4261 ACAGTTTCTAACTGTTCGAATCCATGAGCTATAACTCTGAACGAAAGGGAAAACCTCCAG 4320

Qy      4321 TTTCCCATGCGCAATTCCCAGAGCTCGC 4348
             ||||||||||||||||||||||||||||
Db      4321 TTTCCCATGCGCAATTCCCAGAGCTCGC 4348


RESULT 3
ADX53995
ID   ADX53995 standard; cDNA; 2841 BP.
XX
AC   ADX53995;
XX
DT   21-APR-2005  (first entry)
XX
DE   Plant full length insert polynucleotide seqid 28735.
XX
KW   plant protectant; plant growth regulant; gene therapy; plant;
KW   recombinant DNA construct; physical array; plant breeding marker;
KW   cold tolerance; heat tolerance; drought tolerance; herbicide tolerance;
KW   extreme osmotic condition; pathogen tolerance; pest tolerance;
KW   growth rate; cell cycle pathway; disease resistance;
KW   galactomannan production; lignin production; plant growth regulator;
KW   yield; plant growth; plant development; seed oil; protein yield;
KW   protein content; gene; ss.
XX
OS   Unidentified.
XX
PN   US2004034888-A1.
XX
PD   19-FEB-2004.
XX
PF   28-APR-2003; 2003US-00425114.
XX
PR   06-MAY-1999;   99US-00304517.
PR   05-NOV-2001; 2001US-00985678.
XX
PA   (LIUJ/) LIU J.
PA   (ZHOU/) ZHOU Y.
PA   (KOVA/) KOVALIC D K.
PA   (SCRE/) SCREEN S E.
PA   (TABA/) TABASKA J E.
PA   (CAOY/) CAO Y.
XX
PI   Liu J,  Zhou Y,  Kovalic DK,  Screen SE,  Tabaska JE,  Cao Y;
XX
DR   WPI; 2004-180133/17.
XX
PT   New recombinant DNA construct, useful for improving plant tolerance to
PT   cold, heat, drought, herbicides, extreme osmotic conditions, pathogens or
PT   pests, for conferring increased resistance to plant disease, or for
PT   improving yield.
XX
PS   Claim 1; SEQ ID NO 28735; 15pp; English.
XX
CC   The invention describes a recombinant DNA construct comprising a
CC   polynucleotide consisting of a sequence encoding an amino acid sequence
CC   available in electronic form from the US patent office at
CC   ftp.seqdata.uspto.gov/sequence.html?DocID:2004034888. The polynucleotide
CC   of the invention are also useful in physical arrays of molecules and as
CC   plant breeding markers. The recombinant DNA construct is useful for
CC   improving plant tolerance to cold, heat, drought, herbicides, extreme
CC   osmotic conditions, pathogens or pests, for manipulating growth rate in
CC   plant cells by modification of the cell cycle pathway, for conferring
CC   increased resistance to plant disease, for producing galactomannan,
CC   lignin or plant growth regulators, for increasing the rate of homologous
CC   recombination in plants, for improving yield by modification of
```

```
CC   photosynthesis or carbohydrate, nitrogen or phosphorus use and/or uptake
CC   or by providing improved plant growth and development under at least one
CC   stress condition or for modifying seed oil or protein yield and/or
CC   content. This sequence represents a plant full length insert
CC   polynucleotide that can be used in the recombinant DNA construct of the
CC   invention.
XX
SQ   Sequence 2841 BP; 707 A; 722 C; 730 G; 682 T; 0 U; 0 Other;

  Query Match            20.2%;  Score 880.2;  DB 13;  Length 2841;
  Best Local Similarity  65.3%;  Pred. No. 7.4e-275;
  Matches 1334;  Conservative    0;  Mismatches  688;  Indels   21;  Gaps    2;

Qy       2217 GAGGATACTCTAAATATCGACATGTCAGACATTTATGACCATGATCCTGATCTATACGCA 2276
Db        625 GGGGAGTCGCTCGACGTCGATGCGCACGACGTGTTTGACCACGACCCAGACCTCTACAGC 684

Qy       2277 AAAATTGTTCGATACCCACTCGACATCATCCCCCTGTTGGACACTGAGTGTCAGGAAGTT 2336
Db        685 AAGATGGTTCGCTATCCGCTCGAGGTGCTCGCCATCTTCGACATCGTGCTCATGGACCTC 744

Qy       2337 GCTACCTCTTTACTACCAACGTTTGAGAAGCATATTGAGGCCAGACCTTTCAATCTCAAA 2396
Db        745 GTCGCGCGCATCGAGCCGCTCTTCGAGAAGCACATCCAGACCAGGATCTACAACCTCAAG 804

Qy       2397 GCATCGGTGCACATGCGTGAACTCAACCCTTCAGATATAGACAAATTGGTTTCTGTTAAA 2456
Db        805 TCGTCCATTTGCTTGAGGAATCTCAACCCATCTGATATTGAGAAGATGGTATCCATCAAG 864

Qy       2457 GGAATGGTTATCCGGTGCAGTTCTATCATACCTGAAATTAAGGGGGCCTTCTTCAAATGT 2516
Db        865 GGTATGATAATTAGATGCAGCTCGGTCATACCGGAGCTCAAGGAGGCTGTGTTCCGCTGC 924

Qy       2517 TTAGTGTGTGGTCACTCGCCTCCGCTAGTTACAGTTGTTAAAGGGCGGGTTGAGGAGCCA 2576
Db        925 CTGGTTTGTGGTTTCTACTCAGAGCCCGTCATGGTTGATAGAGGAAGAGTAACTGAACCA 984

Qy       2577 ACAAGGTGTGAAAAGCCAGAATGTGCAGCACGGAATGCTATGTCTCTTATTCACAATCGA 2636
Db        985 CACATTTGTCAGAAAGAACAATGTAAAGCCACAAATTCTATGACCCTAGTGCACAACAGA 1044

Qy       2637 TGCACTTTTGCAAATAAGCAGATAGTGCGTCTTCAAGAAACTCCAGATGCCATTCCTGAA 2696
Db       1045 TGCAGATTTTCAGACAAGCAGATCATAAAGTTGCAGGAAACACCAGACGAGATACCAGAA 1104

Qy       2697 GGAGAGACTCCACACACAGTCAGCATGTGTTTATACAACACTATGGTTGATGCTGTGAAG 2756
Db       1105 GGTGGCACTCCACATACAGTTAGTGTCTTGATGCATGATAAGCTTGTTGATGCTGGAAAG 1164

Qy       2757 CCTGGAGATCGTATTGAGGTAACAGGAGTTTTCAAGGCCATGGCAGTTCGAGTTGGTCCG 2816
Db       1165 CCTGGAGATAGGGTTGAGATAACTGGAATATACAGAGCTATGAGTATTCGAATTGGACCA 1224

Qy       2817 AATCAACGAACATTACGAGCATTGTATAAGACCTACATCGATTGCGTGCACGTCAAGAAG 2876
Db       1225 ACTCAAAGGACAGTGAAGTCTATATTCAAGACATATATTGATTGCCTTCACATAAAGAAG 1284

Qy       2877 TCTGACAGGGGTCGACTGCAAACTGAAGATCCTATGGAGATGGATAAGGAGAATGATATG 2936
Db       1285 ACAGACAAGTCTAGGCTTCATGTGGAGGACACCATGGATATTGATAATTCTAACGCTAGC 1344

Qy       2937 TATGCTGGGTATCATGAAAGTGATACTTCAGAAGCTGCTAATGAAGCAAAGATTCAAAAA 2996
Db       1345 AAATC-----------------TACTGAAGAGGATTTTCTTAGTGATAAGGTTGAGAAA 1386

Qy       2997 CTTAAAGAGCTGTCCAAGCTCCCGGGCATTTATGATAGACTTTCAAGGTCGCTGGCTCCA 3056
Db       1387 CTAAAAGAGCTTTCGAAGTTGCCTGATATCTATGAAAGATTGACTAGATCATTAGCTCCA 1446

Qy       3057 AGCATTTGGGAGCTTGAAGATATTAAAAAGGGTCTTCTTTGCCAGCTCTTTGGTGGAAG 3116
Db       1447 AACATATGGGAGTTGGATGATGTCAAAAGAGGTCTCCTTTGCCAGCTTTTCGGCGGTAAT 1506

Qy       3117 GCTAAGAAAATTCCATCTGGAGCATCTTTCCGAGGTGACATCAATGTTTTACTTGTTGGG 3176
Db       1507 CCCTTGAAGCTTCCTTCTGGAGCTAGTTTCCGGGGTGACATCAATATTTTACTTGTGGGG 1566

Qy       3177 GACCCTGGTACCAGTAAATCTCAGCTGCTTCAGTATGTGCACAAGATAGCTCCTCGTGGA 3236
Db       1567 GACCCTGGAACAAGTAAATCCCAGCTTCTCCAGTACATGCATAAACTGTCTCCTCGTGGT 1626

Qy       3237 ATCTACACTAGTGGGCGAGGAAGTTCGGCGGTTGGGCTGACAGCGTATGTAACGAAGGAT 3296
Db       1627 ATCTATACGAGTGGTAGAGGAAGTTCTGCTGTTGGTCTTACTGCTTATGTTACCAAAGAC 1686

Qy       3297 CCAGAAACTCGAGAGACGGTATTGGAGAGCGGAGCTTTGGTTCTTAGTGATCGTGGGATA 3356
Db       1687 CCTGAGACTGGCGAAACTGTTCTAGAAAGTGGAGCACTTGTTTTGAGTGACAAAGGTGTT 1746

Qy       3357 TGCTGTATCGATGAGTTCGACAAAATGTCTGATAATGCCCGAAGCATGCTTCATGAGGTA 3416
Db       1747 TGTTGCATAGATGAGTTTGATAAGATGTCTGATAATGCCCGAAGCATGTTACACGAGGTG 1806

Qy       3417 ATGGAGCAACAAACGGTATCTGTAGCCAAAGGGGGTATCATTGCCTCGCTGAACGCTCGG 3476
Db       1807 ATGGAACAGCAGACAGTATCCATTGCGAAGGCTGGAATAATTGCATCTTTAAACGCTAGG 1866

Qy       3477 ACGTCTGTCCTTGCATGTGCAAATCCTAGTGGGTCCCGATACAATGCGCGCCTTTCTGTG 3536
Db       1867 ACATCTGTCCTGGCATGTGCCAATCCTACTGAATCACGTTACAATCCAAGGCTCTCTGTA 1926

Qy       3537 ATTGATAACATCCAGCTTCCTCCAACTCTACTTTCTAGATTTGATTTAATTTACTTAATG 3596
Db       1927 ATTGACAACATCCACTTAGCGCCAACGCTACTTTCAAGATTCGACCTGATTTATCTTATC 1986

Qy       3597 CTCGACAAACCAGACGAGCAAAACGATCGTCGTCTCGCCAGGCATCTCGTGGCTTTACAC 3656
Db       1987 TTGGACAAGGCGGATGAGCAAACTGATAGGCGCCTGGCAAAGCATATTGTTTCGTTGCAT 2046

Qy       3657 TATGAAAACTATGAAGTTTCAAAGCAGGACGCCTTAGATCTACAAACACTTACCGCGTAT 3716
Db       2047 TTTGAGAATCCAAATTTAGAGGAGCTCGAGGTCTTGGACTTGCAGACACTAGTTTCCTAC 2106

Qy       3717 ATCACCTATGCTCGTCAGCATGTACATCCTACATTAAGTGATGAAGCTGCTGAAGATTTG 3776
Db       2107 ATAAGCTATGCAAGGAAGTATATTCAGCCACAGTTATCTGATGAAGCTGCAGAAGAGTTA 2166

Qy       3777 ATTAATGGCTATGTTGAGATGCGCCAAAAGGGCAACTTTCCTGGAAGCAGTAAAAAGGTG 3836
Db       2167 ACTCGTGGCTATGTCGAGATGAGAAAAAGAGGGAATAGCCCTGGGAGCAGAAAGAAGGTC 2226

Qy       3837 ATAACAGCCACACCTCGGCAACTCGAAAGTATGATTCGTATCAGTGAAGCCCTAGCTCGA 3896
Db       2227 ATAACAGCAACCGCTAGACAAATAGAGAGTTTGATCCGTCTCAGCGAAGCATTAGCCCGA 2286

Qy       3897 ATGAGATTTTCTGAAGTGGTAGAGAAAGTTGATGCAGCAGAAGCTGTGCGCCTTTTAGAC 3956
Db       2287 ATGCGGTTCTCTGAAGTGGTCGAGGTGCGGGATGTTGTGGAGGCATTCAGGCTTCTTGAA 2346

Qy       3957 GTCGCTTTGCAGCAATCTGCTACTGATCATGCAACAGGTACGATAGACATGGATCTTATC 4016
Db       2347 GTCGCCATGCAGCAGTCTGCGACGGATCATGCAACTGGTACGATTGATATGGATCTGATC 2406

Qy       4017 ACGACTGGAGTGTCGGCCAGCGAGCGTATTCGTCGGGCCAACTTGCTAGCTGCTCTGCGA 4076
Db       2407 ATGACGGGGATATCCGCAAGCGAAAGGCAGAGGCGGGAGAACCTCGTTGCCGCAACCCGT 2466

Qy       4077 GAGCTTATAGCAGATAAAATTTCACCTGGCAGCTCCTCTGGCTTGAAGACCAGTCAGCTT 4136
Db       2467 AACCTGATTGCGGAGAAAATGCAGCTTGGAGGCCCCTC---GATGCGCATGATTGAGTTG 2523

Qy       4137 CTTGAGGATATCCGGAGCCAAAGCAGTGTGGACGTTAGTTTGCAGGATATTAAAAATGCT 4196
Db       2524 CTGGAGGAACTGAGGAAGCAGAGCTCAATGGAAATTCATATGCACGAACTCCGCGGTGCT 2583

Qy       4197 CTGGGTAGCCTCCAAGGAGAAGGCTTTCTTACTGTCCATGGTGACATAGTCAAGAGAGTT 4256
```

```
                || || | |||       |||||     |    | ||||||| ||||  || | |||||||
Db       2584 CTTGGCACCCTGATGACTGAAGGCGCGGTGGTTATCCATGGAGACAACGTGAGGAGAGTT 2643

Qy       4257 TGA 4259
              |||
Db       2644 TGA 2646


RESULT 4
AAI67610
ID   AAI67610 standard; cDNA; 723 BP.
XX
AC   AAI67610;
XX
DT   27-FEB-2002  (first entry)
XX
DE   Cell cycle protein 2 (CC-2) partial cDNA sequence.
XX
KW   Cell Cycle Stress-Related Protein; CCSRP; cell cycle protein; CC-1; CC-2;
KW   CC-3; environmental stress; ss.
XX
OS   Physcomitrella patens.
XX
PN   WO200177354-A2.
XX
PD   18-OCT-2001.
XX
PF   06-APR-2001; 2001WO-US011294.
XX
PR   07-APR-2000; 2000US-0196001P.
XX
PA   (BADI ) BASF PLANT SCI GMBH.
XX
PI   Costa E SilvaO,  Bohnert HJ,  Van Thielen N,  Chen R;
PI   Sarria-Millan R;
XX
DR   WPI; 2002-049151/06.
XX
PT   Novel Cell Cycle Stress-Related Protein useful for increasing tolerance
PT   to environmental stress, is selected from Cell Cycle Proteins 1-3, or
PT   their orthologs.
XX
PS   Disclosure; Fig 1B; 90pp; English.
XX
CC   The invention relates to a Cell Cycle Stress-Related Protein (CCSRP),
CC   isolated from Physcomitrella patens, and selected from Cell Cycle (CC)-1
CC   protein, CC-2 protein, CC-3 protein, or their orthologs. The CCSRP and
CC   encoding nucleic acids are useful for increasing tolerance to
CC   environmental stress selected from salinity, drought and temperature, in
CC   transgenic plants including monocot and dicot selected from maize, wheat,
CC   rye, oat, triticale, rice, barley, soybean, peanut, cotton, rapeseed,
CC   canola, manihot, pepper, sunflower, tagetes, solanaceous plants, potato,
CC   tobacco, eggplant, tomato, Vicia species, pea, alfalfa, coffee, cacao,
CC   tea, Salix species, oil palm, coconut, perennial grass and forage crops.
CC   The nucleic acid is also useful for identifying organisms e.g. P. patens
CC   in a mixed population of microorganisms. The nucleic acids are also
CC   useful for evolutionary and protein structural studies. The proteins and
CC   nucleic acids are useful as markers for specific regions of the genome.
CC   The present sequence represents a partial cDNA sequence of CC-2 from P.
CC   patens
XX
SQ   Sequence 723 BP; 204 A; 136 C; 201 G; 182 T; 0 U; 0 Other;

  Query Match             15.5%;  Score 673;  DB 6;  Length 723;
  Best Local Similarity   97.1%;  Pred. No. 1.1e-207;
  Matches  696;  Conservative    0;  Mismatches   20;  Indels    1;  Gaps    1;

Qy       2753 GAAGCCTGGAGATCGTATTGAGGTAACAGGAGTTTTCAAGGCCATGGCAGTTCGAGTTGG 2812
              | | |  |||||||||||||||||||||||||||||||||||||||||||||||||||||
Db          3 GCACCAGGGAGATCGTATTGAGGTAACAGGAGTTTTCAAGGCCATGGCAGTTCGAGTTGG 62

Qy       2813 TCCGAATCAACGAACATTACGAGCATTGTATAAGACCTACATCGATTGCGTGCACGTCAA 2872
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db         63 TCCGAATCAACGAACATTACGAGCATTGTATAAGACCTACATCGATTGCGTGCACGTCAA 122

Qy       2873 GAAGTCTGACAGGGGTCGACTGCAAACTGAAGATCCTATGGAGATGGATAAGGAGAATGA 2932
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        123 GAAGTCTGACAGGGGTCGACTGCAAACTGAAGATCCTATGGAGATGGATAAGGAGAATGA 182

Qy       2933 TATGTATGCTGGGTATCATGAAAGTGATACTTCAGAAGCTGCTAATGAAGCAAAGATTCA 2992
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        183 TATGTATGCTGGGTATCATGAAAGTGATACTTCAGAAGCTGCTAATGAAGCAAAGATTCA 242

Qy       2993 AAAACTTAAAGAGCTGTCCAAGCTCCCGGGCATTTATGATAGACTTTCAAGGTCGCTGGC 3052
              ||||||||||||||||||||||||||||| ||||||||||||||||||||||||||||||
Db        243 AAAACTTAAAGAGCTGTCCAAGCTCCCGGACATTTATGATAGACTTTCAAGGTCGCTGGC 302

Qy       3053 TCCAAGCATTTGGGAGCTTGAAGATATTAAAAAGGGTCTTCTTTGCCAGCTCTTTGGTGG 3112
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        303 TCCAAGCATTTGGGAGCTTGAAGATATTAAAAAGGGTCTTCTTTGCCAGCTCTTTGGTGG 362

Qy       3113 GAAGGCTAAGAAAATTCCATCTGGAGCATCTTTCCGAGGTGACATCAATGTTTTACTTGT 3172
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        363 GAAGGCTAAGAAAATTCCATCTGGAGCATCTTTCCGAGGTGACATCAATGTTTTACTTGT 422

Qy       3173 TGGGGACCCTGGTACCAGTAAATCTCAGCTGCTTCAGTATGTGCACAAGATAGCTCCTCG 3232
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        423 TGGGGACCCTGGTACCAGTAAATCTCAGCTGCTTCAGTATGTGCACAAGATAGCTCCTCG 482

Qy       3233 TGGAATCTACACTAGTGGGCGAGGAAGTTCGGCGGTTGGGCTGACAGCGTATGT-AACGA 3291
              |||||||||||||||||||||||||||||||||||||||||||||||||||||| |||||
Db        483 TGGAATCTACACTAGTGGGCGAGGAAGTTCGGCGGTTGGGCTGACAGCGTATGTAAACGA 542

Qy       3292 AGGATCCAGAAACTCGAGAGACGGTATTGGAGAGCGGAGCTTTGGTTCTTAGTGATCGTG 3351
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        543 AGGATCCAGAAACTCGAGAGACGGTATTGGAGAGCGGAGCTTTGGTTCTTAGTGATCGTG 602

Qy       3352 GGATATGCTGTATCGATGAGTTCGACAAAATGTCTGATAATGCCCGAAGCATGCTTCATG 3411
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        603 GGATATGCTGTATCGATGAGTTCGACAAAATGTCTGATAATGCCCGAAGCATGCTTCATG 662

Qy       3412 AGGTAATGGAGCAACAAACGGTATCTGTAGCCAAAGGGGGTATCATTGCCTCGCTGA 3468
              |||||||||||||||||||||||||| | ||||  ||    ||  | |    || ||
Db        663 AGGTAATGGAGCAACAAACGGTATCTGGACCCAAGCGGTTCATGCTCGTGAAGCCGA 719


RESULT 5
AAD60810
ID   AAD60810 standard; cDNA; 723 BP.
XX
AC   AAD60810;
XX
DT   15-JAN-2004  (first entry)
XX
DE   Physcomitrella patens CC-2 EST (expressed sequence tag) cDNA.
XX
KW   Cell cycle stress-related protein; CCSRP; metal; cell cycle protein; CC;
KW   transgenic plant; environmental stress; stress tolerance; salinity;
KW   drought; temperature; chemical; oxidative stress; expressed sequence tag;
KW   EST; ss.
XX
OS   Physcomitrella patens.
XX
PN   US2003097675-A1.
XX
PD   22-MAY-2003.
XX
PF   06-APR-2001; 2001US-00828062.
XX
PR   07-APR-2000; 2000US-0196001P.
XX
PA   (SILV/) COSTA E SILVA O.
PA   (BOHN/) BOHNERT H J.
PA   (THIE/) THIELEN N V.
PA   (CHEN/) CHEN R.
PA   (SARR/) SARRIA-MILLAN R.
```

```
XX
PI   Costa E SilvaO,  Bohnert HJ,  Thielen NV,  Chen R,  Sarria-Millan R;
XX
DR   WPI; 2003-765533/72.
XX
PT   Novel cell cycle stress-related protein capable of conferring stress
PT   tolerance such as tolerance towards salinity, drought, temperature,
PT   chemical, pathogens, to plants upon over-expression.
XX
PS   Example 5; Fig 1B; 62pp; English.
XX
CC   The invention relates to an isolated cell cycle stress-related protein
CC   (CCSRP) which is chosen from a cell cycle-1 protein (CC-1), CC-2, CC-3
CC   and its orthologues. Recombinant expression vector is useful for
CC   producing a transgenic plant containing a CCSRP coding nucleic acid,
CC   where expression of the nucleic acid in the plant results in increased
CC   tolerance to environmental stress as compared to a wild type variety of
CC   the plant which involves transforming a plant cell with the recombinant
CC   expression vector, generating from the plant cell a transgenic plant with
CC   an increased tolerance to environmental stress as compared to a wild type
CC   variety of the plant. CCSRP is useful for conferring stress tolerance
CC   such as tolerance towards salinity, drought, temperature, metal,
CC   chemical, pathogens and oxidative stress or their combinations to plants.
CC   The present sequence is Physcomitrella patens CC-2 EST (expressed
CC   sequence tag) cDNA
XX
SQ   Sequence 723 BP; 204 A; 136 C; 201 G; 182 T; 0 U; 0 Other;

  Query Match             15.5%;  Score 673;  DB 10;  Length 723;
  Best Local Similarity   97.1%;  Pred. No. 1.1e-207;
  Matches  696;  Conservative    0;  Mismatches   20;  Indels    1;  Gaps    1;

Qy   2753 GAAGCCTGGAGATCGTATTGAGGTAACAGGAGTTTTCAAGGCCATGGCAGTTCGAGTTGG 2812
          | | |  |||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      3 GCACCAGGGAGATCGTATTGAGGTAACAGGAGTTTTCAAGGCCATGGCAGTTCGAGTTGG 62

Qy   2813 TCCGAATCAACGAACATTACGAGCATTGTATAAGACCTACATCGATTGCGTGCACGTCAA 2872
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     63 TCCGAATCAACGAACATTACGAGCATTGTATAAGACCTACATCGATTGCGTGCACGTCAA 122

Qy   2873 GAAGTCTGACAGGGGTCGACTGCAAACTGAAGATCCTATGGAGATGGATAAGGAGAATGA 2932
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    123 GAAGTCTGACAGGGGTCGACTGCAAACTGAAGATCCTATGGAGATGGATAAGGAGAATGA 182

Qy   2933 TATGTATGCTGGGTATCATGAAAGTGATACTTCAGAAGCTGCTAATGAAGCAAAGATTCA 2992
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    183 TATGTATGCTGGGTATCATGAAAGTGATACTTCAGAAGCTGCTAATGAAGCAAAGATTCA 242

Qy   2993 AAAACTTAAAGAGCTGTCCAAGCTCCCGGGCATTTATGATAGACTTTCAAGGTCGCTGGC 3052
          ||||||||||||||||||||||||||||| ||||||||||||||||||||||||||||||
Db    243 AAAACTTAAAGAGCTGTCCAAGCTCCCGGACATTTATGATAGACTTTCAAGGTCGCTGGC 302

Qy   3053 TCCAAGCATTTGGGAGCTTGAAGATATTAAAAAGGGTCTTCTTTGCCAGCTCTTTGGTGG 3112
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    303 TCCAAGCATTTGGGAGCTTGAAGATATTAAAAAGGGTCTTCTTTGCCAGCTCTTTGGTGG 362

Qy   3113 GAAGGCTAAGAAAATTCCATCTGGAGCATCTTTCCGAGGTGACATCAATGTTTTACTTGT 3172
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    363 GAAGGCTAAGAAAATTCCATCTGGAGCATCTTTCCGAGGTGACATCAATGTTTTACTTGT 422

Qy   3173 TGGGGACCCTGGTACCAGTAAATCTCAGCTGCTTCAGTATGTGCACAAGATAGCTCCTCG 3232
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    423 TGGGGACCCTGGTACCAGTAAATCTCAGCTGCTTCAGTATGTGCACAAGATAGCTCCTCG 482

Qy   3233 TGGAATCTACACTAGTGGGCGAGGAAGTTCGGCGGTTGGGCTGACAGCGTATGT-AACGA 3291
          |||||||||||||||||||||||||||||||||||||||||||||||||||||| |||||
Db    483 TGGAATCTACACTAGTGGGCGAGGAAGTTCGGCGGTTGGGCTGACAGCGTATGTAAACGA 542

Qy   3292 AGGATCCAGAAACTCGAGAGACGGTATTGGAGAGCGGAGCTTTGGTTCTTAGTGATCGTG 3351
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    543 AGGATCCAGAAACTCGAGAGACGGTATTGGAGAGCGGAGCTTTGGTTCTTAGTGATCGTG 602

Qy   3352 GGATATGCTGTATCGATGAGTTCGACAAAATGTCTGATAATGCCCGAAGCATGCTTCATG 3411
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    603 GGATATGCTGTATCGATGAGTTCGACAAAATGTCTGATAATGCCCGAAGCATGCTTCATG 662

Qy   3412 AGGTAATGGAGCAACAAACGGTATCTGTAGCCAAAGGGGGTATCATTGCCTCGCTGA 3468
          ||||||||||||||||||||||||||| | ||||  ||   ||  | |    || ||
Db    663 AGGTAATGGAGCAACAAACGGTATCTGGACCCAAGCGGTTCATGCTCGTGAAGCCGA 719


RESULT 6
ADX36902
ID   ADX36902 standard; cDNA; 1571 BP.
XX
AC   ADX36902;
XX
DT   21-APR-2005  (first entry)
XX
DE   Plant full length insert polynucleotide seqid 19722.
XX
KW   plant protectant; plant growth regulant; gene therapy; plant;
KW   recombinant DNA construct; physical array; plant breeding marker;
KW   cold tolerance; heat tolerance; drought tolerance; herbicide tolerance;
KW   extreme osmotic condition; pathogen tolerance; pest tolerance;
KW   growth rate; cell cycle pathway; disease resistance;
KW   galactomannan production; lignin production; plant growth regulator;
KW   yield; plant growth; plant development; seed oil; protein yield;
KW   protein content; gene; ss.
XX
OS   Unidentified.
XX
PN   US2004034888-A1.
XX
PD   19-FEB-2004.
XX
PF   28-APR-2003; 2003US-00425114.
XX
PR   06-MAY-1999;   99US-00304517.
PR   05-NOV-2001; 2001US-00985678.
XX
PA   (LIUJ/) LIU J.
PA   (ZHOU/) ZHOU Y.
PA   (KOVA/) KOVALIC D K.
PA   (SCRE/) SCREEN S E.
PA   (TABA/) TABASKA J E.
PA   (CAOY/) CAO Y.
XX
PI   Liu J,  Zhou Y,  Kovalic DK,  Screen SE,  Tabaska JE,  Cao Y;
XX
DR   WPI; 2004-180133/17.
XX
PT   New recombinant DNA construct, useful for improving plant tolerance to
PT   cold, heat, drought, herbicides, extreme osmotic conditions, pathogens or
PT   pests, for conferring increased resistance to plant disease, or for
PT   improving yield.
XX
PS   Claim 1; SEQ ID NO 19722; 15pp; English.
XX
CC   The invention describes a recombinant DNA construct comprising a
CC   polynucleotide consisting of a sequence encoding an amino acid sequence
CC   available in electronic form from the US patent office at
CC   ftp.seqdata.uspto.gov/sequence.html?DocID:2004034888. The polynucleotide
CC   of the invention are also useful in physical arrays of molecules and as
CC   plant breeding markers. The recombinant DNA construct is useful for
CC   improving plant tolerance to cold, heat, drought, herbicides, extreme
CC   osmotic conditions, pathogens or pests, for manipulating growth rate in
CC   plant cells by modification of the cell cycle pathway, for conferring
CC   increased resistance to plant disease, for producing galactomannan,
CC   lignin or plant growth regulators, for increasing the rate of homologous
CC   recombination in plants, for improving yield by modification of
CC   photosynthesis or carbohydrate, nitrogen or phosphorus use and/or uptake
CC   or by providing improved plant growth and development under at least one
CC   stress condition or for modifying seed oil or protein yield and/or
CC   content. This sequence represents a plant full length insert
CC   polynucleotide that can be used in the recombinant DNA construct of the
```

```
CC   invention.
XX
SQ   Sequence 1571 BP; 437 A; 311 C; 403 G; 420 T; 0 U; 0 Other;

  Query Match            14.6%;  Score 634.2;  DB 13;  Length 1571;
  Best Local Similarity  67.5%;  Pred. No. 8.3e-195;
  Matches  907;  Conservative    0;  Mismatches  433;  Indels    3;  Gaps    1;

Qy       2917 TGGATAAGGAGAATGATATGTATGCTGGGTATCATGAAAGTGATACTTCAGAAGCTGCTA 2976
Db         32 TGGAGGACACCATGGATATTGATAATTCTAACGCTAGCAAATCTACTGAAGAGGATTTTC 91

Qy       2977 ATGAAGCAAAGATTCAAAAACTTAAAGAGCTGTCCAAGCTCCCGGGCATTTATGATAGAC 3036
Db         92 TTAGTGATAAGGTTGAGAAACTAAAAGAGCTTTCGAAGTTGCCTGATATCTATGAAAGAT 151

Qy       3037 TTTCAAGGTCGCTGGCTCCAAGCATTTGGGAGCTTGAAGATATTAAAAAGGGTCTTCTTT 3096
Db        152 TGACTAGATCATTAGCTCCAAACATATGGGAGTTGGATGATGTCAAAAGAGGTCTCCTTT 211

Qy       3097 GCCAGCTCTTTGGTGGGAAGGCTAAGAAAATTCCATCTGGAGCATCTTTCCGAGGTGACA 3156
Db        212 GCCAGCTTTTCGGCGGCAATCCCTTGAAGCTTCCTTCTGGAGCTAGTTTCCGGGGTGACA 271

Qy       3157 TCAATGTTTTACTTGTTGGGGACCCTGGTACCAGTAAATCTCAGCTGCTTCAGTATGTGC 3216
Db        272 TCAATATTTTACTTGTGGGGGACCCTGGAACAAGTAAATCCCAGCTTCTCCAGTACATGC 331

Qy       3217 ACAAGATAGCTCCTCGTGGAATCTACACTAGTGGGCGAGGAAGTTCGGCGGTTGGCTGA 3276
Db        332 ATAAACTGTCTCCTCGTGGTATCTATACGAGTGGTAGAGGAAGTTCTGCTGTTGGTCTTA 391

Qy       3277 CAGCGTATGTAACGAAGGATCCAGAAACTCGAGAGACGGTATTGGAGAGCGGAGCTTTGG 3336
Db        392 CTGCTTATGTTACCAAAGACCCTGAGACTGGCGAAACTGTTCTAGAAAGTGGAGCACTTG 451

Qy       3337 TTCTTAGTGATCGTGGGATATGCTGTATCGATGAGTTCGACAAAATGTCTGATAATGCCC 3396
Db        452 TTTTGAGTGACAAAGGTGTTTGTTGCGTAGATGAGTTTGATAAGATGTCTGATAATGCCC 511

Qy       3397 GAAGCATGCTTCATGAGGTAATGGAGCAACAAACGGTATCTGTAGCCAAAGGGGGTATCA 3456
Db        512 GAAGCATGTTACACGAGGTGATGGAACAGCAGACAGTATCCATTGCGAAGGCTGGAATAA 571

Qy       3457 TTGCCTCGCTGAACGCTCGGACGTCTGTCCTTGCATGTGCAAATCCTAGTGGGTCCCGAT 3516
Db        572 TTGCATCTTTAAACGCTAGGACATCTGTCCTGGCATGTGCCAATCCTACTGAATCACGTT 631

Qy       3517 ACAATGCGCGCCTTTCTGTGATTGATAACATCCAGCTTCCTCCAACTCTACTTTCTAGAT 3576
Db        632 ACAATCCAAGGCTCTCTGTAATTGACAACATCCACTTAGCGCCAACGCTACTTTCAAGAT 691

Qy       3577 TTGATTTAATTTACTTAATGCTCGACAAACCAGACGAGCAAAACGATCGTCGTCTCGCCA 3636
Db        692 TCGACCTGATTTATCTTATCTTGGACAAGGCGGATGAGCAAACTGATAGGCGCCTGGCAA 751

Qy       3637 GGCATCTCGTGGCTTTACACTATGAAAACTATGAAGTTTCAAAGCAGGACGCCTTAGATC 3696
Db        752 AGCATATTGTTTCGTTGCATTTTGAGAATCCAAATTTAGAGGAGCTCGAGGTCTTGGACT 811

Qy       3697 TACAAACACTTACCGCGTATATCACCTATGCTCGTCAGCATGTACATCCTACATTAAGTG 3756
Db        812 TGCAGACACTAGTTTCCTACATAAGCTATGCAAGGAAGTATATTCAGCCACAGTTATCTG 871

Qy       3757 ATGAAGCTGCTGAAGATTTGATTAATGGCTATGTTGAGATGCGCCAAAAGGGCAACTTTC 3816
Db        872 ATGAAGCTGCAGAAGAGTTAACTCGTGGCTATGTCGAGATGAGAAAAAGAGGGAATAGCC 931

Qy       3817 CTGGAAGCAGTAAAAAGGTGATAACAGCCACACCTCGGCAACTCGAAAGTATGATTCGTA 3876
Db        932 CTGGGAGCAGAAAGAAGGTCATAACAGCAACCGCTAGACAAATAGAGAGTTTGATCCGTC 991

Qy       3877 TCAGTGAAGCCCTAGCTCGAATGAGATTTTCTGAAGTGGTAGAGAAAGTTGATGCAGCAG 3936
Db        992 TCAGCGAAGCATTAGCCCGAATGCGGTTCTCTGAAGTGGTCGAGGTGCGGGATGTTGTGG 1051

Qy       3937 AAGCTGTGCGCCTTTTAGACGTCGCTTTGCAGCAATCTGCTACTGATCATGCAACAGGTA 3996
Db       1052 AGGCATTCAGGCTTCTTGAAGTCGCCATGCAGCAGTCTGCGACGGATCATGCAACTGGTA 1111

Qy       3997 CGATAGACATGGATCTTATCACGACTGGAGTGTCGGCCAGCGAGCGTATTCGTCGGGCCA 4056
Db       1112 CGATTGATATGGATCTGATCATGACGGGGATATCCGCAAGCGAAAGGCAGAGGCGGGAGA 1171

Qy       4057 ACTTGCTAGCTGCTCTGCGAGAGCTTATAGCAGATAAAATTTCACCTGGCAGCTCCTCTG 4116
Db       1172 ACCTCGTTGCCGCAACCCGTAACCTGATTGCGGAGAAAATGCAGCTTGGAGGCCCCTC-- 1229

Qy       4117 GCTTGAAGACCAGTCAGCTTCTTGAGGATATCCGGAGCCAAAGCAGTGTGGACGTTAGTT 4176
Db       1230 -GATGCGCATGATTGAGTTGCTGGAGGAACTGAGGAAGCAGAGCTCAATGGAAATTCATA 1288

Qy       4177 TGCAGGATATTAAAAATGCTCTGGGTAGCCTCCAAGGAGAAGGCTTTCTTACTGTCCATG 4236
Db       1289 TGCACGAACTCCGCGGTGCTCTTGGCACCCTGATGACTGAAGGCGCGGTGGTTATCCATG 1348

Qy       4237 GTGACATAGTCAAGAGAGTTTGA 4259
Db       1349 GAGACAACGTGAGGAGAGTTTGA 1371


RESULT 7
ADB69927
ID   ADB69927 standard; DNA; 2970 BP.
XX
AC   ADB69927;
XX
DT   04-DEC-2003  (first entry)
XX
DE   C. neoformans open reading frame SEQ ID NO:2332.
XX
KW   ds; gene; fungicide; gene therapy; infection.
XX
OS   Cryptococcus neoformans.
XX
PN   WO2003052076-A2.
XX
PD   26-JUN-2003.
XX
PF   17-DEC-2002; 2002WO-US040225.
XX
PR   17-DEC-2001; 2001US-0341261P.
XX
PA   (ELIT-) ELITRA PHARM INC.
XX
PI   Zamudio C,  Eroshkin AM;
XX
DR   WPI; 2003-533017/50.
DR   P-PSDB; ADB70288.
XX
PT   New nucleic acid, useful for preparing a composition for treating an
PT   infection caused by Cryptococcus neoformans.
XX
PS   Claim 2; SEQ ID NO 2332; 136pp; English.
XX
CC   The invention relates to a novel purified or isolated Cryptococcus
CC   neoformans nucleic acid molecule comprising a sequence encoding a
CC   polypeptide comprising a sequence not given in the specification. A
CC   polynucleotide of the invention has fungicide activity, and may have a
CC   use in gene therapy. The nucleic acid is useful for preparing a
CC   composition for treating an infection caused by Cryptococcus neoformans.
CC   The present sequence represents a C. neoformans sequence of the
CC   invention. Note: The sequence data for this patent is not represented in
CC   the printed specification, but was obtained in electronic format directly
CC   from WIPO at ftp.wipo.int/pub/published_pat_sequences.
XX
```

```
SQ   Sequence 2970 BP; 711 A; 746 C; 742 G; 771 T; 0 U; 0 Other;

  Query Match             11.3%;  Score 492.8;  DB 10;  Length 2970;
  Best Local Similarity   57.1%;  Pred. No. 1.7e-148;
  Matches  993;  Conservative    0;  Mismatches  717;  Indels   30;  Gaps    4;

Qy       2372 TGAGGCCAGACCTTTCAATCTCAAAGCATCGGTGCACATGCGTGAACTCAACCCTTCAGA 2431
Db        768 TAAAGTCAGACCGTTTGGTGGAGAGAAGACGGTCAACATGAGGGATCTTAATCCCGGAGA 827

Qy       2432 TATAGACAAATTGGTTTCTGTTAAAGGAATGGTTATCCGGTGCAGTTCTATCATACCTGA 2491
Db        828 TACTGACAAGCTGGTGACCGTGAAGGGTCTTGTCATCCGTGCAACTCCCGTCATTCCTGA 887

Qy       2492 AATTAAGGGGGCCTTCTTCAAATGTTTAGTGTGTGGTCACTCGCCTCCGCTAGTTACAGT 2551
Db        888 TATGACCACTGCTTTCTTCCGATGCTTGGTTTGTCAGCACACTGTTCAAGCGGATA---T 944

Qy       2552 TGTTAAAGGGCGGGTTGAGGAGCCAACAAGGTGTGAAAAGCCAGAATGTGCAGCACGGAA 2611
Db        945 CGACCGAGGAAGAATCAGCGAGCCAGAGCGATGTCCCCGAGACGTCTGTGGTTCAACTGG 1004

Qy       2612 TGCTATGTCTCTTATTCACAATCGATGCACTTTTGCAAATAAGCAGATAGTGCGTCTTCA 2671
Db       1005 CACTATGTCTCTTATTCATAACCGCTCCGAATTCACCGACAAACAAGTCATCCGTCTTCA 1064

Qy       2672 AGAAACTCCAGATGCCATTCCTGAAGGAGAGACTCCACACACAGTCAGCATGTGTTTATA 2731
Db       1065 GGAGACTCCCGATGCTGTACCCGATGGGCAGACTCCACATACTGTATCTCTTTGCGTCTA 1124

Qy       2732 CAACACTATGGTTGATGCTGTGAAGCCTGGAGATCGTATTGAGGTAACAGGAGTTTTCAA 2791
Db       1125 TGATGAACTCGTTGACTTGGTCAAGCCTGGTGACCGAGTGATCATCACTGGTATCTTCCG 1184

Qy       2792 GGCCATGGCAGTTCGAGTTGGTCCGAATCAACGAACATTACGAGCATTGTATAAGACCTA 2851
Db       1185 ATCAATCCCAGTTCGTGTCAACCCTCGCCAAAGGAGTATCAAATCTTTGTACAAGACCTA 1244

Qy       2852 CATCGATTGCGTGCACGTCAAGAAGTCTGACAGGGGTCGACTGCAAACTGAAGATCCTAT 2911
Db       1245 TCTCGATGTCGTCCACGTCAAGCGTACCAACACTGCCCGAATGGGCTTTGACCCTTCTA- 1303

Qy       2912 GGAGATGGATAAGGAGAATGATATGTATGCTGGGTATCATGAAAGTGATACTTCAGAAGC 2971
Db       1304 --------CTCGTGCGGGCGAGAGCAAGCCACCGGGTGACCTTGGGTCGCCTGTCAGATC 1355

Qy       2972 TGCTAATGAAGCAAAGATTCAAAAACTTAAAGAGCTGTCCAAGCTCCCGGGCATTTATGA 3031
Db       1356 AGCCGCTGCGGAGATGGAACAACGCATTATCGAGCTCTCCAACCACCCTGATCTTTACAA 1415

Qy       3032 TAGACTTTCAAGGTCGCTGGCTCCAAGCATTTGGAGCTTGAAGATATTAAAAAGGGTCT 3091
Db       1416 CATACTTGCCTCATCTTTGGCGCCCTCCATCTACGAGCTCGAAGATGTTAAGAAGGGCAT 1475

Qy       3092 TCTTTGCCAGCTCTTTGGTGGGAAGGCTAAGAAAATTCCATCTGGAG------------- 3138
Db       1476 CCTCCTCCAGCTTTTCGGCGGTACCAACAAGTCTATCGCCCGTGGTGGCGGTGGGGGCGG 1535

Qy       3139 --CATCTTTCCGAGGTGACATCAATGTTTTACTTGTTGGGGACCCTGGTACCAGTAAATC 3196
Db       1536 TCCTAGATACAGAGGTGATATCAACGTGTTGATGGTTGGTGATCCTGGTACTAGTAAATC 1595

Qy       3197 TCAGCTGCTTCAGTATGTGCACAAGATAGCTCCTCGTGGAATCTACACTAGTGGGCGAGG 3256
Db       1596 TCAGATTCTGCAGTATGTGCATAAGATTGCTCCTCGAGGTGTCTACACCTCCGGTAAGGG 1655

Qy       3257 AAGTTCGGCGGTTGGGCTGACAGCGTATGTAACGAAGGATCCAGAAACTCGAGAGACGGT 3316
Db       1656 AAGTTCAGCGGTTGGTTTGACGGCCTATGTAACGAGAGATCCCGATTCCAAGCAGCTCGT 1715

Qy       3317 ATTGGAGAGCGGAGCTTTGGTTCTTAGTGATCGTGGGATATGCTGTATCGATGAGTTCGA 3376
Db       1716 TTTGGAAAGTGGTGCATTGGTTTTGTCTGACGGCGGAGTTTGCTGTATCGATGAGTTTGA 1775

Qy       3377 CAAAATGTCTGATAATGCCCGAAGCATGCTTCATGAGGTAATGGAGCAACAAACGGTATC 3436
Db       1776 CAAGATGAGTGATGCCACTCGAAGTGTCTTGCACGAAGTAATGGAGCAACAAACGGTTTC 1835

Qy       3437 TGTAGCCAAAGGGGGTATCATTGCCTCGCTGAACGCTCGGACGTCTGTCCTTGCATGTGC 3496
Db       1836 TATTGCCAAGGCAGGTATCATCACCACTCTCAATGCTCGTACATCCATTCTCGCCGCCGC 1895

Qy       3497 AAATCCTAGTGGGTCCCGATACAATGCGCGCCTTTCTGTGATTGATAACATCCAGCTTCC 3556
Db       1896 GAACCCAATCAACTCTCGATATGACCCTAACCTCCCTATTCCTGCCAACATCGATCTCCC 1955

Qy       3557 TCCAACTCTACTTTCTAGATTTGATTTAATTTACTTAATGCTCGACAAACCAGACGAGCA 3616
Db       1956 TCCTACATTGATCTCAAGATTCGACTTGCTGTACTTGGTGTTGGACAAGGTGGATGAGGT 2015

Qy       3617 AAACGATCGTCGTCTCGCCAGGCATCTCGTGGCTTTACACTATGAAAACTATGAAGTTTC 3676
Db       2016 GAACGACACAGAAAGTTGGCAAAGCATTTGGTGGGACTCTATTTGTCAGATGTAGAGGATCA 2075

Qy       3677 AAAGCAGGAC---GCCTTAGATCTACAAACACTTACCGCGTATATCACCTATGCTCGTCA 3733
Db       2076 GCCTGCCGACAATATCATTCCTCTTCAGACCTTAACGTCCTACATTACATATGCCCGCTC 2135

Qy       3734 GCATGTACATCCTACATTAAGTGATGAAGCTGCTGAAGATTTGATTAATGGCTATGTTGA 3793
Db       2136 CAAAATCCACCCCGTCCTCACCGAAGGAGCTTCTGAAGCGCTTGTTCAGGCCTATGTTGA 2195

Qy       3794 GATGCGCCAAAAGGGCAACTTTCCTGGAAGCAGTAAAAAGGTGATAACAGCCACACCTCG 3853
Db       2196 GATGCGCAAGGCCGGTATGGACTCTCGTACTCAGGAAAAGAGGATCACTGCTACCACTAG 2255

Qy       3854 GCAACTCGAAAGTATGATTCGTATCAGTGAAGCCCTAGCTCGAATGAGATTTTCTGAAGT 3913
Db       2256 GCAGTTGGAAAGTATGATCAGATTGGGCGAAGCGCATGCGAGGATGAGATTAAGTGACAG 2315

Qy       3914 GGTAGAGAAAGTTGATGCAGCAGAAGCTGTGCGCCTTTTAGACGTCGCTTTGCAGCAATC 3973
Db       2316 GGTTGAGGAAGAGGACATTAGGGAGGCTGTTAGGTTGATCAAGAGTGCCCTGAGGGAGAG 2375

Qy       3974 TGCTACTGATCATGCAACAGGTACGATAGACATGGATCTTATCACGACTGGAGTGTCGGC 4033
Db       2376 TGCTACCGATCCCCTTACTGGTCAAATTGATCTTGACCTCATCAACACCGGTGCCGGCCA 2435

Qy       4034 CAGCGAGCGTATTCGTCGGGCCAACTTGCTAGCTGCTCTGCGAGAGCTTATAGCAGATAA 4093
Db       2436 AACCATGCGTCGCGCCCGAGCCGATCTTAAACGTGAAGTTATCAAACTCGTCGTCGAAAA 2495

RESULT 8
ADJ39580
ID   ADJ39580 standard; cDNA; 1723 BP.
XX
AC   ADJ39580;
XX
DT   06-MAY-2004  (first entry)
XX
DE   Plant cDNA #580.
XX
KW   Plant; gene; ss; transcription; plant genome augmentation; cereal;
KW   soybean; alfalfa; sunflower; canola; cotton; peanut; tobacco; sugar beet;
KW   maize; barley; sorghum; rice; wheat; crop plant; insecticide resistance;
KW   stress tolerance; salt tolerance; cold tolerance; drought tolerance;
KW   plant nutrition; apical dominance; dwarfism; early flowering; antiviral;
KW   antifungal.
XX
OS   Eukaryota.
XX
PN   US2004016025-A1.
XX
PD   22-JAN-2004.
XX
```

```
PF   26-SEP-2002; 2002US-00260238.
XX
PR   26-SEP-2001; 2001US-0325277P.
PR   26-SEP-2001; 2001US-0325448P.
PR   04-APR-2002; 2002US-0370620P.
XX
PA   (BUDW/) BUDWORTH P.
PA   (MOUG/) MOUGHAMER T.
PA   (BRIG/) BRIGGS S P.
PA   (COOP/) COOPER B.
PA   (GLAZ/) GLAZEBROOK J.
PA   (GOFF/) GOFF S A.
PA   (KATA/) KATAGIRI F.
PA   (KREP/) KREPS J.
PA   (PROV/) PROVART N.
PA   (RICK/) RICKE D.
PA   (ZHUT/) ZHU T.
XX
PI   Budworth P,  Moughamer T,  Briggs SP,  Cooper B,  Glazebrook J;
PI   Goff SA,  Katagiri F,  Kreps J,  Provart N,  Ricke D,  Zhu T;
XX
DR   WPI; 2004-190374/18.
XX
PT   New rice promoter, useful for manipulating crop plants to alter or
PT   improve phenotypic characteristics, e.g. produce large quantities of oil
PT   or proteins, resistance to insecticides, virus or fungi, stress tolerance
PT   or high nutritional value.
XX
PS   Claim 70; SEQ ID NO 580; 230pp; English.
XX
CC   The invention relates to plant nucleotide sequences that direct seed-,
CC   leaf- and/or stem-, panicle-, root- or pollen-specific or -preferential
CC   or constitutive transcription of an operatively linked nucleic acid
CC   segment. The invention also relates to a method for augmenting a plant
CC   genome and a method of identifying a gene, where its expression is
CC   altered in the seed, leaf, stem, panicle, pollen, root or is constitutive
CC   in a plant cell. The plant is a cereal, e.g. soybean, alfalfa, sunflower,
CC   canola, cotton, peanut, tobacco or sugar beet, preferably maize, barley,
CC   sorghum, rice or wheat. The polynucleotides and the polypeptides they
CC   encode are useful for manipulating crop plants to alter or improve
CC   phenotypic characteristics, to produce large quantities of oil or
CC   proteins, to incur resistance to insecticides, viruses or fungi, and to
CC   incur stress tolerance (e.g. salt, cold or drought) to ensure the plants
CC   have a high nutritional value with reduced apical dominance or dwarfism,
CC   early flowering or altered metabolic pathways. This sequence represents a
CC   plant nucleic acid of the invention. Note: The sequence data for this
CC   patent did not form part of the printed specification but was obtained in
CC   electronic format directly from USPTO at seqdata.uspto.gov/sequence.html.
XX
SQ   Sequence 1723 BP; 459 A; 373 C; 461 G; 429 T; 0 U; 1 Other;

  Query Match             11.3%;  Score 490;  DB 12;  Length 1723;
  Best Local Similarity   69.4%;  Pred. No. 9.4e-148;
  Matches  762;  Conservative    0;  Mismatches  285;  Indels   51;  Gaps    5;

Qy       2989 TTCAAAAACTTAAAGAGCTGTCCAAGCTCCCGGGCATTTATGATAGACTTTCAAGGTCGC 3048

Db        606 TTCACATAAAGAAGGAGTTGTCAAAGTTGCCAGATATATATGACAGATTAACTAGGTCAT 665

Qy       3049 TGGCTCCAAGCATTTGGGAGCTTGAAGATATTAAAAAGGGTCTTCTTTGC---CAGCTCT 3105

Db        666 TGGCTCCAAACATATGGGAGCTGGACGACGTTAAAAGGGGCCTCCTTTGCCAGCAGCTTT 725

Qy       3106 TTGGTGGGAAGGCTAAGAAAATTCCATCTGGAGCATCTTTCCGAGGTGACATCAATGTTT 3165

Db        726 TTGGTGGAAATGCTTTGAGGCTTCCTTCTGGAGCTAGTTTCCGAGGCGACATCAATATTT 785

Qy       3166 TACTTGTTGGGGACCCTGGTACCAGTAAATCTCAGCTGCTTCAGTATGTGCACAAGATAG 3225

Db        786 TGCTTGTTGGTGATCCTGGAACAAGTAAATCCCAGCTTCTCCAATACATGCACAAACTGT 845

Qy       3226 CTCCTCGTGGAATCTACACTAGTGGGCGAGGAAGTTCGGCGGTTGGGCTGACAGCGTATG 3285

Db        846 CTCCTCGTGGCATTTATACAAGTGGCAGAGGAAGTTCAGCTGTTGGCCTTACTGCATACG 905

Qy       3286 TAACGAAGGATCCAGAAACTCGAGAGACGGTATTGGAGAGCGGAGCTTTGGTTCTTAGTG 3345

Db        906 TTACCAAGGATCCTGAAACTGGTGAAACTGTTCTTGAGAGTGGAGCGCTTGTTTTGAGTG 965

Qy       3346 ATCGTGGGATATGCTGTATCGATGAGTTCGACAAAATGTCTGATAATGCCCGAAGCATGC 3405

Db        966 ACAAAGGTGTTTGTTGTATTGATGAATTTGATAAGATGTCTGATAATGCTCGAAGCATGT 1025

Qy       3406 TTCATGAGGTAATGGAGCAACAAACGGTATCTGTAGCCAAAGGGGGTATCATTGCCTCGC 3465

Db       1026 TACATGAGGTGATGGAACAACAGACTGTCTCCATTGCCAAGGCTGGAATAATTGCATCTT 1085

Qy       3466 TGAACGCTCGGACGTCTGTCCTTGCATGTGCAAATCCTAGTGGGTCCCGATACAATGCGC 3525

Db       1086 TAAATGCCAGAACATCAGTTCTAGCATGTGCAAATCCTACTGAATCACGTTATAATCCAA 1145

Qy       3526 GCCTTTCTGTGATTGATAACATCCAGCTTCCTCCAACTCTACTTTCTAGA--TTTGATTT 3583

Db       1146 GGCTCTCTGTGATTGACAATATCCATCTTCCTCCAACACTGCTTTCTAGAGGTTTGACCT 1205

Qy       3584 AATTTACTTAATGCTCGACAAACCAGACGAGCAAAACGATCGTCGTCTCGCCAGGCATCT 3643

Db       1206 CATTTATCTGATATTGGACAAGGCAGATGAGCAAACTGATAGACGCCTGGCTAAGCATAT 1265

Qy       3644 CGTGGCTTTACACTATGAAAACTATGAAGTTTCAAAGCAGGACGCCTTAGATCTACAAAC 3703

Db       1266 TGTTTCGTTGCATTTTGAGAA--------------------------------------- 1286

Qy       3704 ACTTACCGCGTATATCACCTATGCTCGTCAGCATGTACATCCTACATTAAGTGATGAAGC 3763

Db       1287 ---TCCAAACTACATAAGTTATGCAAGGAAGCATATACAACCACAGTTATCTGATGAAGC 1343

Qy       3764 TGCTGAAGATTTGATTAATGGCTATGTTGAGATGCGCCAAAAGGGCAACTTTCCTGGAAG 3823

Db       1344 TGCAGAAGAATTGACTCGCGGCTATGTTGAGATGAGGAAAAGAGGAAACAGCCCTGGTAG 1403

Qy       3824 CAGTAAAAAGGTGATAACAGCCACACCTCGGCAACTCGAAAGTATGATTCGTATCAGTGA 3883

Db       1404 CAG---AAAGAAGATAACTGCGACAGCTCGACAAATTGAGAGCTTGATTCGGCTCAGTGA 1460

Qy       3884 AGCCCTAGCTCGAATGAGATTTTCTGAAGTGGTAGAGAAAGTTGATGCAGCAGAAGCTGT 3943

Db       1461 AGCACTGGCCCGAATGCGATTCTCTGAAATGGTTGAAGTACAAGATGTTGTAGAGGCCTT 1520

Qy       3944 GCGCCTTTTAGACGTCGCTTTGCAGCAATCTGCTACTGATCATGCAAC-AGGTACGATAG 4002

Db       1521 CAGGCTTCTCGAAGTTGCCATGCAGCAATCGGCAACTGATCATGCCACTGGGTACAATCG 1580

Qy       4003 ACATGGATCTTATCACGACTGGAGTGTCGGCCAGCGAGCGTATTCGTCGGGCCAACTTGC 4062

Db       1581 ATATGGATCTTATCATGACTGGAATATCTGCGAGCGAAAGGCAGAGGCGGGACCAGTTGC 1640

Qy       4063 TAGCTGCTCTGCGAGAGC 4080

Db       1641 TGGAAGAAATTAGGAAGC 1658


RESULT 9
ABZ32193
ID   ABZ32193 standard; DNA; 2733 BP.
XX
AC   ABZ32193;
XX
DT   30-JAN-2003  (first entry)
XX
DE   Candida albicans essential gene SEQ ID NO 6480.
XX
KW   Fungus; yeast; tetracyclin; promoter; GRACE strain; biosynthesis;
KW   signal transduction; DNA replication; cell division; growth;
KW   proliferation; Candida albicans; fungicide; antifungal; gene; ss.
XX
```

```
OS   Candida albicans.
XX
PN   WO200253728-A2.
XX
PD   11-JUL-2002.
XX
PF   26-DEC-2001; 2001WO-US049486.
XX
PR   29-DEC-2000; 2000US-0259128P.
PR   20-FEB-2001; 2001US-00792024.
PR   22-AUG-2001; 2001US-0314050P.
XX
PA   (ELIT-) ELITRA PHARM INC.
XX
PI   Roemer T,  Jiang B,  Boone C,  Bussey H,  Ohlsen KL;
XX
DR   WPI; 2002-566694/60.
DR   P-PSDB; ABP73643.
XX
PT   Constructing strains for identifying gene products as effective targets
PT   for therapeutic intervention, by inactivating in the strain one allele of
PT   a gene and placing other allele of the gene under conditional expression.
XX
PS   Claim 37; SEQ ID NO 6480; 167pp + Sequence Listing; English.
XX
CC   The invention relates to constructing (M1) a strain of diploid fungal
CC   cells in which both alleles of a gene are modified, comprising modifying
CC   one allele by insertion or replacement by a cassette having an
CC   expressible selectable marker and modifying other allele by
CC   recombination, of a promoter replacement fragment with a heterologous
CC   promoter, so that expression of the second allele is regulated by the
CC   promoter. (M1) is useful for constructing a strain of diploid fungal
CC   cells in which both alleles of a gene are modified. The diploid fungal
CC   cells having both alleles modified are useful for identifying a gene that
CC   is essential to the survival or growth of a fungus, a gene that
CC   contributes to the virulence and/or pathogenicity of a fungus, a gene
CC   that contributes to the resistance of a diploid fungus to an antifungal
CC   agent, an antifungal agent that inhibits the growth of a diploid fungus
CC   and for identifying a therapeutic agent for treatment of a mammalian
CC   disease. (M1) is useful for identifying a compound which modulates the
CC   activity of a gene product, preferably enzymatic activity, carbon
CC   compound catabolism, biosynthetic, transporter, transcriptional,
CC   translational, signal transduction, DNA replication and cell division
CC   activity. The method is useful for identifying a compound having the
CC   ability to inhibit growth or proliferation of C. albicans cells and for
CC   treating infection by C. albicans. The present sequence is that of an
CC   essential Candida albicans gene used in the method of the invention.
CC   Note: The sequence data for this patent is not represented in the printed
CC   specification but is based on sequence information supplied to Derwent by
CC   the European Patent Office
XX
SQ   Sequence 2733 BP; 956 A; 409 C; 508 G; 860 T; 0 U; 0 Other;

  Query Match             11.0%;  Score 480.2;  DB 6;  Length 2733;
  Best Local Similarity   56.6%;  Pred. No. 2e-144;
  Matches  959;  Conservative    0;  Mismatches  718;  Indels   18;  Gaps    3;

Qy       2409 ATGCGTGAACTCAACCCTTCAGATATAGACAAATTGGTTTCTGTTAAAGGAATGGTTATC 2468
Db        883 ATTCGTGAATTGAATCCTAATGATATTGATAAGTTAGTTAGTGTTAAAGGGTTAACATTA 942

Qy       2469 CGGTGCAGTTCTATCATACCTGAAATTAAGGGGGCCTTCTTCAAATGTTTAGTGTGTGGT 2528
Db        943 CGTTCTACTTCTATCATTCCTGATATGAAAGTGGCATTTTTCAGATGTAACGCCTGTGGT 1002

Qy       2529 CACTCGCCTCCGCTAGTTACAGTTGTTAAAGGGCGGGTTGAGGAGCCAACAAGGTGTGAA 2588
Db       1003 CATACTGTTGGG---GTGGAAATTGATCGTGGTGTAATTTCCGAACCAACAAAATGTCCT 1059

Qy       2589 AAGCCAGAATGTGCAGCACGGAATGCTATGTCTCTTATTCACAATCGATGCACTTTTGCA 2648
Db       1060 AGAGAAGTATGTGGTCAAACCAATTCAATGGTACTTATTCATAATCGTTCATCATTTTCT 1119

Qy       2649 AATAAGCAGATAGTGCGTCTTCAAGAAACTCCAGATGCCATTCCTGAAGGAGAGACTCCA 2708
Db       1120 GATAAACAAGTGATTAAATTACAAGAAACTCCTGATTTGGTACCTGATGGTCAAACCCCT 1179

Qy       2709 CACACAGTCAGCATGTGTTTATACAACACTATGGTTGATGCTGTGAAGCCTGGAGATCGT 2768
Db       1180 CATTCTATCAATTTATGTGTTTACGATGAATTGGTTGATTCTTGTCGTGCTGGTGATCGT 1239

Qy       2769 ATTGAGGTAACAGGAGTTTTCAAGGCCATGGCAGTTCGAGTTGGTCCGAATCAACGAACA 2828
Db       1240 GTTGAAGTATGTGGTATTTTCAGATCAACTCCAGTTCGAGCAAATCCTAGACAAAGAGCT 1299

Qy       2829 TTACGAGCATTGTATAAGACCTACATCGATTGCGTGCACGTCAAGAAGTCTGACAGGGGT 2888
Db       1300 TTAAAAAACTTATATAAAACATATTTGGATATTGTTCATGTGAAAAAAATCGATAAAAGG 1359

Qy       2889 CGACTGCAAACTGAAGATCCTATGGAGATGGATAAGGAGAATGATATGTATGCTGGGTAT 2948
Db       1360 CGATTAGGTGGTGATGTTACCA---------CTTTAGAACACGAATTGGCTGAAAAGGAT 1410

Qy       2949 CATGAAAGTGATACTTCAGAAGCTGCTAATGAAGCAAAGATTCAAAAACTTAAAGAGCTG 3008
Db       1411 CAAGAAGTTGAACAAGTTCGAAAAATCACTGCTGAAGAAGAAGCTAAAATCAAAGAAATT 1470

Qy       3009 TCCAAGCTCCCGGGCATTTATGATAGACTTTCAAGGTCGCTGGCTCCAAGCATTTGGGAG 3068
Db       1471 TCTGAACGTGATGATTTGTATGAAATATTGGCTAGATCATTAGCTCCATCTATTTATGAG 1530

Qy       3069 CTTGAAGATATTAAAAAGGGTCTTCTTTGCCAGCTCTTTGGTGGGAAGGCTAAGAAAATT 3128
Db       1531 ATGGATGATGTTAAAAAGGGGATCTTGTTACAATTATTTGGTGGTACTAATAAAACATTT 1590

Qy       3129 CCATCTGGAGCATCTTTCCGAGGTGACATCAATGTTTTACTTGTTGGGGACCCTGGTACC 3188
Db       1591 ACAAAAGGTGGTCGTTATAGAGGTGATATAAATATTTTATTATGTGGTGATCCATCTACT 1650

Qy       3189 AGTAAATCTCAGCTGCTTCAGTATGTGCACAAGATAGCTCCTCGTGGAATCTACACTAGT 3248
Db       1651 TCTAAATCACAAATCTTACAATATGTTCATAAGATTGCTCCAAGAGGGGTTTATACTTCT 1710

Qy       3249 GGGCGAGGAAGTTCGGCGGTTGGGCTGACAGCGTATGTAACGAAGGATCCAGAAACTCGA 3308
Db       1711 GGTAAAGGTTCTTCGGCTGTTGGTTTGACAGCTTATATCACCAGAGATATTGATACAAAA 1770

Qy       3309 GAGACGGTATTGGAGAGCGGAGCTTTGGTTCTTAGTGATCGTGGGATATGCTGTATCGAT 3368
Db       1771 CAATTAGTATTGGAAAGTGGGGCCCTTGTGTTATCTGATGGAGGTGTTTGTTGTATTGAT 1830

Qy       3369 GAGTTCGACAAAAATGTCTGATAATGCCCGAAGCATGCTTCATGAGGTAATGGAGCAACAA 3428
Db       1831 GAGTTTGACAAGATGAGTGATGCCACCAGATCAGTTTTACATGAAGTCATGGAACAACAA 1890

Qy       3429 ACGGTATCTGTAGCCAAAGGGGGTATCATTGCCTCGCTGAACGCTCGGACGTCTGTCCTT 3488
Db       1891 ACTATTTCTATTGCTAAAGCCGGTATTATTACCACATTAAATGCTAGGACATCTATTTTA 1950

Qy       3489 GCATGTGCAAATCCTAGTGGGTCCCGATACAATGCGCGCCTTTCTGTGATTGATAACATC 3548
Db       1951 GCTTCTGCAAATCCAATTAATTCTCGTTATGATCCAAATTTGCCTGTGACGGGTAACATT 2010

Qy       3549 CAGCTTCCTCCAACTCTACTTTCTAGATTTGATTTAATTTACTTAATGCTCGACAAACCA 3608
Db       2011 GATTTGCCACCACCATTGTTATCACGTTTTGATTTGGTTTATTTGATTTTGGATAAAGTT 2070

Qy       3609 GACGAGCAAAACGATCGTCGTCTCGCCAGGCATCTCGTGGCTTTACACTATGAAAACTAT 3668
Db       2071 GATGAATCAATCGATAGACAATTGGCTAGACATTTGACTGATATGTATTTGGAAGATGCT 2130

Qy       3669 GAAGTTTCAAAGCAGGACGCCTTAGATCTAC------AAACACTTACCGCGTATATCACC 3722
Db       2131 CCTGAAACTGTTAATGCCAATTCGGTATTACCCGTGGAATTATTAACACTTTATATTCAA 2190

Qy       3723 TATGCTCGTCAGCATGTACATCCTACATTAAGTGATGAAGCTGCTGAAGATTTGATTAAT 3782
```

```
                 ||||||   || || |  ||||      | |  || |||| |    | || ||  |||
Db        2191 TATGCTAAAGAGAATTTCAATCCCGTGATGACAGAAGAAGGTAAAAATGAATTAGTTAGA 2250

Qy        3783 GGCTATGTTGAGATGCGCCAAAAGGGCAACTTTCCTGGAAGCAGTAAAAAGGTGATAACA 3842
                   || ||||| ||| |  ||  |||  |    | || ||       |||||   || ||
Db        2251 TCATACGTTGAAATGAGAAAATTGGGGGAGGATGCTAGATCATCAGAAAAGAGAATCACC 2310

Qy        3843 GCCACACCTCGGCAACTCGAAAGTATGATTCGTATCAGTGAAGCCCTAGCTCGAATGAGA 3902
                || ||  |   | |||  | |||    ||||| |  |    |||||| |   |||  |||||||
Db        2311 GCTACTACAAGACAATTGGAATCGATGATTAGATTGTCTGAAGCTCATGCTAAAATGAGA 2370

Qy        3903 TTTTCTGAAGTGGTAGAGAAAGTTGATGCAGCAGAAGCTGTGCGCCTTTTAGACGTCGCT 3962
                || ||||||    || ||       ||||||  ||   |   |   |    |     ||
Db        2371 TTATCTGAAAGAGTTGAATTGATCGATGTTAAAGAAGCAGTAAGATTAATTAAATCGGCC 2430

Qy        3963 TTGCAGCAATCTGCTACTGATCATGCAACAGGTACGATAGACATGGATCTTATCACGACT 4022
                 |  |  | | | ||||||||||   |  || ||||   || ||  |||||| | ||    |||
Db        2431 ATTAAAGATTATGCTACTGATCCAGTTACTGGTAGAATCGATATGGATATGATTCAAACT 2490

Qy        4023 GGAGTGTCGGCCAGCGAGCGTATTCGTCGGGCCAACTTGCTAGCTGCTCTGCGAGAGCTT 4082
                ||     ||||     |  | |      | |   | |||| |   ||    |      |  |
Db        2491 GGTACTACGGCACAACAAAGAAGAGTACAAGAAGATTTGGTCAGTGAAATTATGAAAATC 2550

Qy        4083 ATAGCAGATAAAATT 4097
                || | ||| || | |
Db        2551 ATCGAAGAAAATAAT 2565


RESULT 10
ADT47760
ID   ADT47760 standard; cDNA; 2802 BP.
XX
AC   ADT47760;
XX
DT   02-DEC-2004  (first entry)
XX
DE   Bacterial polynucleotide #22511.
XX
KW   Recombinant DNA construct; transformed plant; improved plant property;
KW   cold tolerance; heat tolerance; drought tolerance; herbicide; osmosis;
KW   pathogen tolerance; pest tolerance; plant disease resistance;
KW   cell cycle pathway modification; plant growth regulator;
KW   homologous recombination; seed oil yield; protein yield; carbohydrate;
KW   nitrogen; phosphorus; photosynthesis; lignin; galactomannan;
KW   bacterial polynucleotide; gene; ss.
XX
OS   Bacteria.
XX
PN   US2003233675-A1.
XX
PD   18-DEC-2003.
XX
PF   20-FEB-2003; 2003US-00369493.
XX
PR   21-FEB-2002; 2002US-0360039P.
XX
PA   (CAOY/) CAO Y.
PA   (HINK/) HINKLE G J.
PA   (SLAT/) SLATER S C.
PA   (CHEN/) CHEN X.
PA   (GOLD/) GOLDMAN B S.
XX
PI   Cao Y,  Hinkle GJ,  Slater SC,  Chen X,  Goldman BS;
XX
DR   WPI; 2004-061375/06.
XX
PT   New recombinant DNA construct comprising a promoter positioned to provide
PT   for expression of a polynucleotide encoding a polypeptide from a
PT   microbial source, useful for producing plants with improved properties.
XX
PS   Claim 1; SEQ ID NO 46198; 122pp; English.
XX
CC   The invention relates to a recombinant DNA construct comprising a
CC   promoter functional in a plant cell, where the promoter is positioned to
CC   provide for expression of a polynucleotide encoding a polypeptide from a
CC   microbial source. The invention also relates to a transformed plant
CC   comprising the recombinant DNA construct and a method of producing a
CC   transformed plant having an improved property. The plant is a crop plant
CC   such as maize or soybean. The method of producing a transformed plant
CC   having an improved property comprises transforming a plant with the
CC   recombinant DNA construct and growing the transformed plant, where the
CC   polynucleotide or polypeptide is useful for improving plant properties.
CC   The recombinant DNA construct is useful for producing plants with
CC   improved plant properties, e.g. improved cold, heat or drought tolerance,
CC   tolerance to herbicides, extreme osmotic conditions, pathogens or pests,
CC   increased resistance to plant disease, better growth rate by modification
CC   of the cell cycle pathway with plant growth regulators, increased rate of
CC   homologous recombination, modified seed oil or protein yield and/or
CC   content, improved yield by modification of carbohydrate, nitrogen or
CC   phosphorus use and/or uptake, by modification of photosynthesis or by
CC   providing improved plant growth and development under at least one stress
CC   condition, improved lignin production or improved galactomannan
CC   production. This sequence represents a bacterial polynucleotide used in
CC   the scope of the invention. Note: The sequence data for this patent did
CC   not form part of the printed specification but was obtained in electronic
CC   format from USPTO at seqdata.uspto.gov/sequence.html.
XX
SQ   Sequence 2802 BP; 887 A; 541 C; 579 G; 795 T; 0 U; 0 Other;

  Query Match             10.8%;  Score 470.8;  DB 13;  Length 2802;
  Best Local Similarity   55.6%;  Pred. No. 2.4e-141;
  Matches  948;  Conservative    0;  Mismatches  747;  Indels    9;  Gaps    2;

Qy        2364 AAGCATATTGAGGCCAGACCTTTCAATCTCAAAGCATCGGTGCACATGCGTGAACTCAAC 2423
                |||  |   |  |  |  || |||| ||||  |        | |        ||||| ||| | ||
Db         895 AAGTTTTATAAAGTAAGGCCTTACAATGTAGGTTCTTGTAAAGGTATGCGCGAATTGAAT 954

Qy        2424 CCTTCAGATATAGACAAATTGGTTTCTGTTAAAGGAATGGTTATCCGGTGCAGTTCTATC 2483
                ||      ||||| || ||||||   |  || ||   |    |   |  |   |  |   |  |
Db         955 CCAAATGATATTGATAAATTGATAAACTTAAAGGGTCTTGTCCTGAGATCAACTCCGGTT 1014

Qy        2484 ATACCTGAAATTAAGGGGGCCTTCTTCAAATGTTTAGTGTGTGGTCACTCGCCTCCGCTA 2543
                || |||||| || ||||   || || ||||||||       || || |  |||        |    |   |
Db        1015 ATCCCTGATATGAAGGTAGCGTTTTTCAAATGCAACGTCTGCGATCA---TACAATGGCA 1071

Qy        2544 GTTACAGTTGTTAAAGGGCGGGTTGAGGAGCCAACAAGGTGTGAAAAGCCAGAATGTGCA 2603
                ||    |  |||  || |||          |   |  |||||   |  |||||||||         || |||
Db        1072 GTGGAAATTGATAGAGGAGTTATACAAGAGCCCGCTAGGTGTGAACGTATTGATTGTAAT 1131

Qy        2604 GCACGGAATGCTATGTCTCTTATTCACAATCGATGCACTTTTGCAAATAAGCAGATAGTG 2663
                | ||    |||  | |||||  || ||||||||   | ||   |  ||  |||  |||| ||  |    |
Db        1132 GAACCCAATTCCATGTCACTGATTCACAACAGGTGTTCATTCGCAGATAAACAAGTCATT 1191

Qy        2664 CGTCTTCAAGAAACTCCAGATGCCATTCCTGAAGGAGAGACTCCACACACAGTCAGCATG 2723
                     |  ||  ||||||||||          ||||| |||  |  ||  ||  |||  |    ||    |
Db        1192 AAGTTACAGGAAACTCCAGACTTTGTGCCTGATGGACAAACGCCTCACTCTATCTCATTA 1251

Qy        2724 TGTTTATACAACACTATGGTTGATGCTGTGAAGCCTGGAGATCGTATTGAGGTAACAGGA 2783
                |||  |  |||  |          |  ||  |||  ||       |  |   |  ||  ||||||||||||  ||  ||
Db        1252 TGTGTTTACGATGAATTAGTGGATTCTTGCAGGGCGGGCGATCGTATTGAGGTGACTGGC 1311

Qy        2784 GTTTTCAAGGCCATGGCAGTTCGAGTTGGTCCGAATCAACGAACATTACGAGCATTGTAT 2843
                    ||||  |  ||||    |    ||  ||| |     |    |   |   |||||     ||  |     ||||||
Db        1312 ACGTTCAGGTCCATCCCCATTAGAGCTAATTCCAGGCAACGCGTACTAAAGTCGTTGTAT 1371

Qy        2844 AAGACCTACATCGATTGCGTGCACGTCAAGAAGTCTGACAGGGGTCGACTGCAAACTGAA 2903
                || ||  ||| |||||     || |||||  ||  ||     |                   |       |   |       ||
Db        1372 AAAACATACGTCGATGTGGTCCACGTTAAAAAAGTTTCAGATAAGAGGTTAGACGTCGAT 1431

Qy        2904 GATCCTATGGAGATGGATAAGGAGAATGATATGTATGCTGGGTATCATGAAAGTGATACT 2963
                  | |||                 |   |        |   |    |   |    |   |     ||  ||    ||
Db        1432 ACTTCTACTATTGAACAAGAATTAATGCAGAACAAGGTAGATCATAACGAGGTCGAAGAA 1491

Qy        2964 TCAGAAGCTGCTAATGAAGCAAAGATTCAAAAACTTAAAGAGCTGTCCAAGCTCCCGGGC 3023
                   |   |        || |||       |   |     ||||  ||    |||  |   |                 |
Db        1492 GTAAGACAAATTACTGATCAGGATTTAGCAAAAATTCGCGAGGTTGCGGCAAGAGAAGAT 1551
```

```
Qy   3024 ATTTATGATAGACTTTCAAGGTCGCTGGCTCCAAGCATTTGGGAGCTTGAAGATATTAAA 3083
Db   1552 TTATACAGTTTATTAGCCCGCTCTATTGCCCCAAGTATTTACGAGCTAGAAGATGTCAAG 1611

Qy   3084 AAGGGTCTTCTTTGCCAGCTCTTTGGTGGGAAGGCTAAGAAAATTCCATCTGGAGCATCT 3143
Db   1612 AAGGGTATATTACTTCAGCTATTTGGCGGCACGAATAAAACCTTTACCAAGGGTGGCCGT 1671

Qy   3144 TTCCGAGGTGACATCAATGTTTTACTTGTTGGGGACCCTGGTACCAGTAAATCTCAGCTG 3203
Db   1672 TATAGAGGTGACATAAATATTTTACTTTGTGGGGATCCTTCTACTTCCAAATCGCAAATT 1731

Qy   3204 CTTCAGTATGTGCACAAGATAGCTCCTCGTGGAATCTACACTAGTGGGCGAGGAAGTTCG 3263
Db   1732 TTGCAATACGTTCACAAAATTACTCCTCGTGGTGTGTATACTTCGGGTAAAGGTTCATCT 1791

Qy   3264 GCGGTTGGGCTGACAGCGTATGTAACGAAGGATCCAGAAACTCGAGAGACGGTATTGGAG 3323
Db   1792 GCCGTTGGTTTAACTGCTTATATTACAAGGGATGTCGACACAAAACAACTTGTTTTGGAA 1851

Qy   3324 AGCGGAGCTTTGGTTCTTAGTGATCGTGGGATATGCTGTATCGATGAGTTCGACAAAATG 3383
Db   1852 AGTGGTGCATTAGTATTGTCTGATGGAGGTGTTTGTTGTATTGACGAGTTTGATAAAATG 1911

Qy   3384 TCTGATAATGCCCGAAGCATGCTTCATGAGGTAATGGAGCAACAAACGGTATCTGTAGCC 3443
Db   1912 AGTGATTCTACAAGGTCCGTCTTGCACGAAGTCATGGAACAGCAGACTATTTCAATCGCA 1971

Qy   3444 AAAGGGGGTATCATTGCCTCGCTGAACGCTCGGACGTCTGTCCTTGCATGTGCAAATCCT 3503
Db   1972 AAAGCGGGAATTATCACAACACTTAATGCCAGAAGTTCTATTTTGGCCAGTGCTAACCCA 2031

Qy   3504 AGTGGGTCCCGATACAATGCGCGCCTTTCTGTGATTGATAACATCCAGCTTCCTCCAACT 3563
Db   2032 ATTGGTTCACGCTACAACCCCAATTTGCCTGTGACTGAAAATATTGATCTACCGCCCCCA 2091

Qy   3564 CTACTTTCTAGATTTGATTTAATTTACTTAATGCTCGACAAACCAGACGAGCAAAACGAT 3623
Db   2092 CTACTTTCGAGATTCGATCTGGTCTATCTTGTCCTTGATAAGGTTGATGAGAAAAATGAC 2151

Qy   3624 CGTCGTCTCGCCAGGCATCTCGTGGCTTTACACTATGAAAA------CTATGAAGTTTCA 3677
Db   2152 AGAGAACTAGCCAAACACTTAACAAATCTTTACCTGGAAGATAAGCCCGAACATATTTCT 2211

Qy   3678 AAGCAGGACGCCTTAGATCTACAAACACTTACCGCGTATATCACCTATGCTCGTCAGCAT 3737
Db   2212 CAAGACGACGTGCTACCAGTTGAATTTTTAACGATGTATATTAGTTATGCAAAGGAGCAC 2271

Qy   3738 GTACATCCTACATTAAGTGATGAAGCTGCTGAAGATTTGATTAATGGCTATGTTGAGATG 3797
Db   2272 ATACACCCAATAATCACCGAGGCCGCTAAGACTGAGCTTGTTCGTGCTTATGTAGGAATG 2331

Qy   3798 CGCCAAAAGGGCAACTTTCCTGGAAGCAGTAAAAAGGTGATAACAGCCACACCTCGGCAA 3857
Db   2332 AGAAAGATGGGTGACGATTCGAGATCCGATGAGAAGAGAATCACAGCTACCACAAGACAA 2391

Qy   3858 CTCGAAAGTATGATTCGTATCAGTGAAGCCCTAGCTCGAATGAGATTTTCTGAAGTGGTA 3917
Db   2392 CTTGAAAGTATGATTCGTTTGGCTGAGGCGCACGCCAAAATGAAATTGAAAAACGTCGTA 2451

Qy   3918 GAGAAAGTTGATGCAGCAGAAGCTGTGCGCCTTTTAGACGTCGCTTTGCAGCAATCTGCT 3977
Db   2452 GAGCTGGAGGATGTTCAAGAAGCCGTTAGATTAATTAGATCAGCCATAAAAGATTATGCA 2511

Qy   3978 ACTGATCATGCAACAGGTACGATAGACATGGATCTTATCACGACTGGAGTGTCGGCCAGC 4037
Db   2512 ACAGACCCTAAAACCGGTAAAATCGACATGAATTTAGTTCAAACAGGTAAATCAGTTATT 2571

Qy   4038 GAGCGTATTCGTCGGGCCAACTTG 4061
Db   2572 CAGAGAAAACTACAGGAGGATTTG 2595
```

```
RESULT 11
ABT19388
ID   ABT19388 standard; DNA; 2625 BP.
XX
AC   ABT19388;
XX
DT   16-APR-2003  (first entry)
XX
DE   Aspergillus fumigatus essential gene #1746.
XX
KW   Fungicide; cytostatic; essential gene; Aspergillus fumigatus; infection;
KW   cancer; contamination; biofilm; antibody; immune response; ds.
XX
OS   Aspergillus fumigatus.
XX
PN   WO200286090-A2.
XX
PD   31-OCT-2002.
XX
PF   23-APR-2002; 2002WO-US013142.
XX
PR   23-APR-2001; 2001US-0285697P.
PR   27-APR-2001; 2001US-0287066P.
PR   05-JUN-2001; 2001US-0295890P.
PR   09-JUL-2001; 2001US-0303899P.
PR   31-AUG-2001; 2001US-0316362P.
XX
PA   (ELIT-) ELITRA PHARM INC.
XX
PI   Jiang B,  Tishkoff D,  Zamudio C,  Broshkin AM,  Hu W,  Lemieux SM;
XX
DR   WPI; 2003-093124/08.
XX
PT   New purified or isolated nucleic acids of essential genes of Aspergillus
PT   fumigatus, useful for treating or preventing infections by A. fumigatus,
PT   or for treating a non-infectious disease in a subject e.g. cancer.
XX
PS   Disclosure; Page; 175pp; English.
XX
CC   The invention relates to novel purified or isolated nucleic acids of
CC   essential genes of Aspergillus fumigatus. The isolated nucleic acids of
CC   the invention are used to treat or prevent infections by a pathogenic
CC   organism such as A. fumigatus, to treat a non-infectious disease in a
CC   subject (e.g. cancer), to prevent or contain contamination of an object
CC   by A. fumigatus, or to prevent or inhibit formation on a surface of a
CC   biofilm comprising A. fumigatus. The polynucleotides are useful for
CC   expressing recombinant protein for characterisation, screening or
CC   therapeutic use, as markers for host tissues in which the pathogenic
CC   organisms invade or reside, for comparing with the DNA sequence of A.
CC   fumigatus to identify duplicated genes or paralogues having the same or
CC   similar biochemical activity and/or function, for comparing with DNA
CC   sequences of other related or distant pathogenic organisms to identify
CC   potential orthologous essential or virulence genes, for selecting and
CC   making oligomers for attachment to a nucleic acid array for examination
CC   of expression patterns, for raising anti-protein antibodies, as an
CC   antigen to raise anti-DNA antibodies or to elicit another immune
CC   response, and for identifying polynucleotides encoding the other protein
CC   with which binding occurs or to identify inhibitors of the binding
CC   interaction. The polypeptides may be used to raise antibodies or to
CC   elicit immune response, as a reagent in assays designed to quantitatively
CC   determine levels of the protein in biological fluids, as a marker for
CC   host tissues in which pathogenic organism invade or reside, and to
CC   isolate correlative receptors or ligands in the case or virulence
CC   factors. This polynucleotide sequence represents one of the essential
CC   genes of Aspergillus fumigatus of the invention
XX
SQ   Sequence 2625 BP; 614 A; 704 C; 713 G; 594 T; 0 U; 0 Other;

  Query Match             10.4%;  Score 451.6;  DB 8;  Length 2625;
  Best Local Similarity   56.1%;  Pred. No. 4.2e-135;
  Matches  944;  Conservative    0;  Mismatches  719;  Indels   21;  Gaps    4;
```

```
Qy   2382 CCTTTCAATCTCAAAGCATCGGTGCACATGCGTGAACTCAACCCTTCAGATATAGACAAA 2441
Db    724 CCTTTTGGTCTCGACTCGACTGTGAATATGAGAGACCTCGATCCTGCAGACATGGATAAA 783

Qy   2442 TTGGTTTCTGTTAAAGGAATGGTTATCCGGTGCAGTTCTATCATACCTGAAATTAAGGGG 2501
Db    784 CTAGTAAGCATTAAGGGCTTAGTCATTCGAACGACACCCATCATTCCTGATATGAAAGAA 843

Qy   2502 GCCTTCTTCAAATGTTTAGTGTGTGGTCACTCGCCTCCGCTAGTTACAGTTGTTAAAGGG 2561
Db    844 GCTTTCTTCCGTTGCCAAGTCTGCAACCATGGTGTTCAGGTCGACA---TTGATCGTGGA 900

Qy   2562 CGGGTTGAGGAGCCAACAAGGTGTGAAAAGCCAGAATGTGCAGCACGGAATGCTATGTCT 2621
Db    901 AAGATTGCGGAACCCACTGAGTGCCCACGTCCAGTGTGTAAGGAACGAAACTCGATGCAA 960

Qy   2622 CTTATTCACAATCGATGCACTTTTGCAAATAAGCAGATAGTGCGTCTTCAAGAAACTCCA 2681
Db    961 CTCATCCATAACCGCTGTGTATTTGCCGACAAGCAGGTCATCAAGTTGCAGGAAACACCT 1020

Qy   2682 GATGCCATTCCTGAAGGAGAGACTCCACACACAGTCAGCATGTGTTTATACAACACTATG 2741
Db   1021 GACAGCATTCCTGATGGCCAGACTCCTCACTCGGTTTCCCTTTGTGTGTATGATGAGCTG 1080

Qy   2742 GTTGATGCTGTGAAGCCTGGAGATCGTATTGAGGTAACAGGAGTTTTCAAGGCCATGGCA 2801
Db   1081 GTGGATGTCTGCAAGGCTGGTGATCGGGTCGAAGTGACCGGTATTTTCCGGTGCAACCCT 1140

Qy   2802 GTTCGAGTTGGTCCGAATCAACGAACATTACGAGCATTGTATAAGACCTACATCGATTGC 2861
Db   1141 GTGCGCGTTAATCCTCGCCAGCGTACACAGAAGTCGCTGTTCAAGACGTACATAGATGTT 1200

Qy   2862 GTGCACGTCAAGAAGTCTGACAGGGGTCGACTGCAAACTGAAGATCCTATGGAGATGGAT 2921
Db   1201 CTTCATGTTCAGAAGATCGATCGCAAGAAGTTGGGTATCGACGTCTCGACCATCGAGCAG 1260

Qy   2922 AAGGAGAATGATATGTATGCTGGGTATCATGAAAGTGATACTTCAGAAGCTGCTAATGAA 2981
Db   1261 GAGCTCTCGGAACAGGCGGCTGGGGATGCAGAACAAACACGTAGGCTCACTGCGGAGGAG 1320

Qy   2982 GCAAAGATTCAAAAACTTAAAGAGCTGTCCAAGCTCCCGGGCATTTATGATAGACTTTCA 3041
Db   1321 GAAGAG------AAAATTAAGCGAACTGCTACCAGACCTGATCTGTATGAGCTTCTCTCT 1374

Qy   3042 AGGTCGCTGGCTCCAAGCATTTGGGAGCTTGAAGATATTAAAAAGGGTCTTCTTTGCCAG 3101
Db   1375 CGGTCCTTGGCCCCCAGCATCTACGAGATGGACGACGTGAAGAAGGGAATCCTGCTTCAG 1434

Qy   3102 CTCTTTGGTGGGAAGGCTAAGAAAATTCCATCTGGAG------CATCTTTCCGAGGTGAC 3155
Db   1435 TTGTTTGGAGGCACCAACAAGACCTTCCAGAAGGGTGGTAACCCACGATACCGTGGAGAT 1494

Qy   3156 ATCAATGTTTTACTTGTTGGGGACCCTGGTACCAGTAAATCTCAGCTGCTTCAGTATGTG 3215
Db   1495 ATCAATATCCTTCTCTGTGGTGACCCATCTACATCCAAGTCCCAGCTTCTTCGTTACGTC 1554

Qy   3216 CACAAGATAGCTCCTCGTGGAATCTACACTAGTGGGCGAGGAAGTTCGGCGGTTGGGCTG 3275
Db   1555 CATAAGATTGCCCCTCGCGGTGTGTATACCAGCGGCAAGGGCTCCTCGGCTGTTGGTCTT 1614

Qy   3276 ACAGCGTATGTAACGAAGGATCCAGAAACTCGAGAGACGGTATTGGAGAGCGGAGCTTTG 3335
Db   1615 ACGGCGTACGTCACCCGCGATCCTGAAACCCGCCAGATGGTCCTCGAGTCGGGTGCCTTG 1674

Qy   3336 GTTCTTAGTGATCGTGGGATATGCTGTATCGATGAGTTCGACAAAATGTCTGATAATGCC 3395
Db   1675 GTTCTTTCAGACGGCGGTATCTGTTGCATCGACGAGTTCGACAAGATGAACGAATCCACT 1734

Qy   3396 CGAAGCATGCTTCATGAGGTAATGGAGCAACAAACGGTATCTGTAGCCAAAGGGGGTATC 3455
Db   1735 CGGTCCGTTCTGCATGAAGTCATGGAACAACAGACAGTATCTATCGCCAAGGCAGGCATT 1794

Qy   3456 ATTGCCTCGCTGAACGCTCGGACGTCTGTCCTTGCATGTGCAAATCCTAGTGGGTCCCGA 3515
Db   1795 ATCACTACTTTGAACGCTAGGACCAGCATCCTGGCTTCCGCCAATCCGATCGGTAGCAGG 1854

Qy   3516 TACAATGCGCGCCTTTCTGTGATTGATAACATCCAGCTTCCTCCAACTCTACTTTCTAGA 3575
Db   1855 TACAATCCCAACTTGCCCGTTCCTCAAAATATTGACCTTCCGCCTACCTTGCTCTCCCGA 1914

Qy   3576 TTTGATTTAATTTACTTAATGCTCGACAAACCAGACGAGCAAAACGATCGTCGTCTCGCC 3635
Db   1915 TTCGACTTGGTATACCTCGTGCTGGACCGAGTGGATGAGCAGGAAGATCGTCGGCTCGCT 1974

Qy   3636 AGGCATCTCGTGGCTTTACACTATGAAAAC------TATGAAGTTTCAAAGCAGGACGCC 3689
Db   1975 AAGCATCTTGTCAATATGTACCTGGAAGACAGACCTGAGCATGCTGCCGAGCAAGAAATC 2034

Qy   3690 TTAGATCTACAAACACTTACCGCGTATATCACCTATGCTCGTCAGCATGTACATCCTACA 3749
Db   2035 TTGCCGATCGAATTCCTTACAGCCTATATCACCTACGCCAAGACCAAAGTCCATCCAGTG 2094

Qy   3750 TTAAGTGATGAAGCTGCTGAAGATTTGATTAATGGCTATGTTGAGATGCGCCAAAAGGGC 3809
Db   2095 CTCACACCGGCCGCCGGTAAAGCCTTGTCGGATGCTTACGTTAACATGCGTAAGCTTGGA 2154

Qy   3810 AACTTTCCTGGAAGCAGTAAAAAGGTGATAACAGCCACACCTCGGCAACTCGAAAGTATG 3869
Db   2155 GATGACATCCGGTCTTCTGACCGCCGTATCACCGCTACCACTCGTCAACTGGAGTCCATG 2214

Qy   3870 ATTCGTATCAGTGAAGCCCTAGCTCGAATGAGATTTTCTGAAGTGGTAGAGAAAGTTGAT 3929
Db   2215 ATCCGACTGTCGGAAGCGCATGCGCGTATGCGGCTATCGCCGGAGGTCACTGCGGATGAT 2274

Qy   3930 GCAGCAGAAGCTGTGCGCCTTTTAGACGTCGCTTTGCAGCAATCTGCTACTGATCATGCA 3989
Db   2275 GTGGAGGAAGCCGTGCGCCTGATCCGCTCCGCCATCAAGCAGGCGGCCACTGACTCTCGG 2334

Qy   3990 ACAGGTACGATAGACATGGATCTTATCACGACTGGAGTGTCGGCCAGCGAGCGTATTCGT 4049
Db   2335 ACCGGTCTGATCGACATGAGCTTGTTGACGGAGGGCACTAGTGCCAGCGAGAGACGCAGC 2394

Qy   4050 CGGG 4053
Db   2395 CGGG 2398
```

```
RESULT 12
ABT21208
ID   ABT21208 standard; DNA; 3072 BP.
XX
AC   ABT21208;
XX
DT   16-APR-2003  (first entry)
XX
DE   Aspergillus fumigatus essential gene #3566.
XX
KW   Fungicide; cytostatic; essential gene; Aspergillus fumigatus; infection;
KW   cancer; contamination; biofilm; antibody; immune response; ds.
XX
OS   Aspergillus fumigatus.
XX
PN   WO200286090-A2.
XX
PD   31-OCT-2002.
XX
PF   23-APR-2002; 2002WO-US013142.
XX
PR   23-APR-2001; 2001US-0285697P.
PR   27-APR-2001; 2001US-0287066P.
PR   05-JUN-2001; 2001US-0295890P.
PR   09-JUL-2001; 2001US-0303899P.
PR   31-AUG-2001; 2001US-0316362P.
XX
PA   (ELIT-) ELITRA PHARM INC.
XX
```

```
PI   Jiang B,  Tishkoff D,  Zamudio C,  Eroshkin AM,  Hu W,  Lemieux SM;
XX
DR   WPI; 2003-093124/08.
XX
PT   New purified or isolated nucleic acids of essential genes of Aspergillus
PT   fumigatus, useful for treating or preventing infections by A. fumigatus,
PT   or for treating a non-infectious disease in a subject e.g. cancer.
XX
PS   Disclosure; Page; 175pp; English.
XX
CC   The invention relates to novel purified or isolated nucleic acids of
CC   essential genes of Aspergillus fumigatus. The isolated nucleic acids of
CC   the invention are used to treat or prevent infections by a pathogenic
CC   organism such as A. fumigatus, to treat a non-infectious disease in a
CC   subject (e.g. cancer), to prevent or contain contamination of an object
CC   by A. fumigatus, or to prevent or inhibit formation on a surface of a
CC   biofilm comprising A. fumigatus. The polynucleotides are useful for
CC   expressing recombinant protein for characterisation, screening or
CC   therapeutic use, as markers for host tissues in which the pathogenic
CC   organisms invade or reside, for comparing with the DNA sequence of A.
CC   fumigatus to identify duplicated genes or paralogues having the same or
CC   similar biochemical activity and/or function, for comparing with DNA
CC   sequences of other related or distant pathogenic organisms to identify
CC   potential orthologous essential or virulence genes, for selecting and
CC   making oligomers for attachment to a nucleic acid array for examination
CC   of expression patterns, for raising anti-protein antibodies, as an
CC   antigen to raise anti-DNA antibodies or to elicit another immune
CC   response, and for identifying polynucleotides encoding the other protein
CC   with which binding occurs or to identify inhibitors of the binding
CC   interaction. The polypeptides may be used to raise antibodies or to
CC   elicit immune response, as a reagent in assays designed to quantitatively
CC   determine levels of the protein in biological fluids, as a marker for
CC   host tissues in which pathogenic organism invade or reside, and to
CC   isolate correlative receptors or ligands in the case or virulence
CC   factors. This polynucleotide sequence represents one of the essential
CC   genes of Aspergillus fumigatus of the invention
XX
SQ   Sequence 3072 BP; 717 A; 847 C; 827 G; 681 T; 0 U; 0 Other;

  Query Match             10.4%;  Score 451.6;  DB 8;  Length 3072;
  Best Local Similarity   56.1%;  Pred. No. 4.7e-135;
  Matches  944;  Conservative    0;  Mismatches  719;  Indels   21;  Gaps    4;

Qy       2382 CCTTTCAATCTCAAAGCATCGGTGCACATGCGTGAACTCAACCCTTCAGATATAGACAAA 2441
Db       1171 CCTTTTGGTCTCGACTCGACTGTGAATATGAGAGACCTCGATCCTGCAGACATGGATAAA 1230

Qy       2442 TTGGTTTCTGTTAAAGGAATGGTTATCCGGTGCAGTTCTATCATACCTGAAATTAAGGGG 2501
Db       1231 CTAGTAAGCATTAAGGGCTTAGTCATTCGAACGACACCCATCATTCCTGATATGAAAGAA 1290

Qy       2502 GCCTTCTTCAAATGTTTAGTGTGTGGTCACTCGCCTCCGCTAGTTACAGTTGTTAAAGGG 2561
Db       1291 GCTTTCTTCCGTTGCCAAGTCTGCAACCATGGTGTTCAGGTCGACA---TTGATCGTGGA 1347

Qy       2562 CGGGTTGAGGAGCCAACAAGGTGTGAAAAGCCAGAATGTGCAGCACGGAATGCTATGTCT 2621
Db       1348 AAGATTGCGGAACCCACTGAGTGCCCACGTCCAGTGTGTAAGGAACGAAACTCGATGCAA 1407

Qy       2622 CTTATTCACAATCGATGCACTTTTGCAAATAAGCAGATAGTGCGTCTTCAAGAAACTCCA 2681
Db       1408 CTCATCCATAACCGCTGTGTATTTGCCGACAAGCAGGTCATCAAGTTGCAGGAAACACCT 1467

Qy       2682 GATGCCATTCCTGAAGGAGAGACTCCACACACAGTCAGCATGTGTTTATACAACACTATG 2741
Db       1468 GACAGCATTCCTGATGGCCAGACTCCTCACTCGGTTTCCCTTTGTGTGTATGATGAGCTG 1527

Qy       2742 GTTGATGCTGTGAAGCCTGGAGATCGTATTGAGGTAACAGGAGTTTTCAAGGCCATGGCA 2801
Db       1528 GTGGATGTCTGCAAGGCTGGTGATCGGGTCGAAGTGACCGGTATTTTCCGGTGCAACCCT 1587

Qy       2802 GTTCGAGTTGGTCCGAATCAACGAACATTACGAGCATTGTATAAGACCTACATCGATTGC 2861
Db       1588 GTGCGCGTTAATCCTCGCCAGCGTACACAGAAGTCGCTGTTCAAGACGTACATAGATGTT 1647

Qy       2862 GTGCACGTCAAGAAGTCTGACAGGGGTCGACTGCAAACTGAAGATCCTATGGAGATGGAT 2921
Db       1648 CTTCATGTTCAGAAGATCGATCGCAAGAAGTTGGGTATCGACGTCTCGACCATCGAGCAG 1707

Qy       2922 AAGGAGAATGATATGTATGCTGGGTATCATGAAAGTGATACTTCAGAAGCTGCTAATGAA 2981
Db       1708 GAGCTCTCGGAACAGGCGGCTGGGGATGCAGAACAAACACGTAGGCTCACTGCGGAGGAG 1767

Qy       2982 GCAAAGATTCAAAAACTTAAAGAGCTGTCCAAGCTCCCGGGCATTTATGATAGACTTTCA 3041
Db       1768 GAAGAG------AAAATTAAGCGAACTGCTACCAGACCTGATCTGTATGAGCTTCTCTCT 1821

Qy       3042 AGGTCGCTGGCTCCAAGCATTTGGGAGCTTGAAGATATTAAAAAGGGTCTTCTTTGCCAG 3101
Db       1822 CGGTCCTTGGCCCCCAGCATCTACGAGATGGACGACGTGAAGAAGGGAATCCTGCTTCAG 1881

Qy       3102 CTCTTTGGTGGGAAGGCTAAGAAAATTCCATCTGGAG------CATCTTTCCGAGGTGAC 3155
Db       1882 TTGTTTGGAGGCACCAACAAGACCTTCCAGAAGGGTGGTAACCCACGATACCGTGGAGAT 1941

Qy       3156 ATCAATGTTTTACTTGTTGGGGACCCTGGTACCAGTAAATCTCAGCTGCTTCAGTATGTG 3215
Db       1942 ATCAATATCCTTCTCTGTGGTGACCCATCTACATCCAAGTCCCAGCTTCTTCGTTACGTC 2001

Qy       3216 CACAAGATAGCTCCTCGTGGAATCTACACTAGTGGGCGAGGAAGTTCGGCGGTTGGGCTG 3275
Db       2002 CATAAGATTGCCCCTCGCGGTGTGTATACCAGCGGCAAGGGCTCCTCGGCTGTTGGTCTT 2061

Qy       3276 ACAGCGTATGTAACGAAGGATCCAGAAACTCGAGAGACGGTATTGGAGAGCGGAGCTTTG 3335
Db       2062 ACGGCGTACGTCACCCGCGATCCTGAAACCCGCCAGATGGTCCTCGAGTCGGGTGCCTTG 2121

Qy       3336 GTTCTTAGTGATCGTGGGATATGCTGTATCGATGAGTTCGACAAAATGTCTGATAATGCC 3395
Db       2122 GTTCTTTCAGACGGCGGTATCTGTTGCATCGACGAGTTCGACAAGATGAACGAATCCACT 2181

Qy       3396 CGAAGCATGCTTCATGAGGTAATGGAGCAACAAACGGTATCTGTAGCCAAAGGGGGTATC 3455
Db       2182 CGGTCCGTTCTGCATGAAGTCATGGAACAACAGACAGTATCTATCGCCAAGGCAGGCATT 2241

Qy       3456 ATTGCCTCGCTGAACGCTCGGACGTCTGTCCTTGCATGTGCAAATCCTAGTGGGTCCCGA 3515
Db       2242 ATCACTACTTTGAACGCTAGGACCAGCATCCTGGCTTCCGCCAATCCGATCGGTAGCAGG 2301

Qy       3516 TACAATGCGCGCCTTTCTGTGATTGATAACATCCAGCTTCCTCCAACTCTACTTTCTAGA 3575
Db       2302 TACAATCCCAACTTGCCCGTTCCTCAAAATATTGACCTTCCGCCTACCTTGCTCTCCCGA 2361

Qy       3576 TTTGATTTAATTTACTTAATGCTCGACAAACCAGACGAGCAAAACGATCGTCGTCTCGCC 3635
Db       2362 TTCGACTTGGTATACCTCGTGCTGGACCGAGTGGATGAGCAGGAAGATCGTCGGCTCGCT 2421

Qy       3636 AGGCATCTCGTGGCTTTACACTATGAAAAC------TATGAAGTTTCAAAGCAGGACGCC 3689
Db       2422 AAGCATCTTGTCAATATGTACCTGGAAGACAGACCTGAGCATGCTGCCGAGCAAGAAATC 2481

Qy       3690 TTAGATCTACAAACACTTACCGCGTATATCACCTATGCTCGTCAGCATGTACATCCTACA 3749
Db       2482 TTGCCGATCGAATTCCTTACAGCCTATATCACCTACGCCAAGACCAAAGTCCATCCAGTG 2541

Qy       3750 TTAAGTGATGAAGCTGCTGAAGATTTGATTAATGGCTATGTTGAGATGCGCCAAAAGGGC 3809
Db       2542 CTCACACCGGCCGCCGGTAAAGCCTTGTCGGATGCTTACGTTAACATGCGTAAGCTTGGA 2601

Qy       3810 AACTTTCCTGGAAGCAGTAAAAAGGTGATAACAGCCACACCTCGGCAACTCGAAAGTATG 3869
Db       2602 GATGACATCCGGTCTTCTGACCGCCGTATCACCGCTACCACTCGTCAACTGGAGTCCATG 2661

Qy       3870 ATTCGTATCAGTGAAGCCCTAGCTCGAATGAGATTTTCTGAAGTGGTAGAGAAAGTTGAT 3929
Db       2662 ATCCGACTGTCGGAAGCGCATGCGCGTATGCGGCTATCGCCGGAGGTCACTGCGGATGAT 2721
```

```
Qy       3930 GCAGCAGAAGCTGTGCGCCTTTTAGACGTCGCTTTGCAGCAATCTGCTACTGATCATGCA 3989
              |  |  ||||| ||||||||  |    |   |||  |  ||||  | || |||||    |
Db       2722 GTGGAGGAAGCCGTGCGCCTGATCCGCTCCGCCATCAAGCAGGCGGCCACTGACTCTCGG 2781

Qy       3990 ACAGGTACGATAGACATGGATCTTATCACGACTGGAGTGTCGGCCAGCGAGCGTATTCGT 4049
              || |||  ||| ||||||     |  | |||   ||        ||||||||| |     |
Db       2782 ACCGGTCTGATCGACATGAGCTTGTTGACGGAGGGCACTAGTGCCAGCGAGAGACGCAGC 2841

Qy       4050 CGGG 4053
              ||||
Db       2842 CGGG 2845


RESULT 13
ABL65258
ID   ABL65258 standard; DNA; 3273 BP.
XX
AC   ABL65258;
XX
DT   15-MAY-2002  (first entry)
XX
DE   Lung cancer related gene sequence SEQ ID NO:3595.
XX
KW   Human; cancer; colon; breast; ovary; oesophagus; kidney; thyroid;
KW   stomach; lung; prostate; pancreas; carcinoma; antitumour; cancerous;
KW   cytostatic; gene therapy; antineoplastic; Wilm's tumour; adenocarcinoma;
KW   gene; ds.
XX
OS   Homo sapiens.
XX
PN   WO200194629-A2.
XX
PD   13-DEC-2001.
XX
PF   30-MAY-2001; 2001WO-US010838.
XX
PR   05-JUN-2000; 2000US-0209473P.
PR   05-JUN-2000; 2000US-0209531P.
PR   18-SEP-2000; 2000US-0233133P.
PR   18-SEP-2000; 2000US-0233617P.
PR   20-SEP-2000; 2000US-0234009P.
PR   20-SEP-2000; 2000US-0234034P.
PR   20-SEP-2000; 2000US-0234052P.
PR   22-SEP-2000; 2000US-0234509P.
PR   22-SEP-2000; 2000US-0234567P.
PR   25-SEP-2000; 2000US-0234923P.
PR   25-SEP-2000; 2000US-0234924P.
PR   25-SEP-2000; 2000US-0235077P.
PR   25-SEP-2000; 2000US-0235082P.
PR   25-SEP-2000; 2000US-0235134P.
PR   25-SEP-2000; 2000US-0235280P.
PR   26-SEP-2000; 2000US-0235637P.
PR   26-SEP-2000; 2000US-0235638P.
PR   27-SEP-2000; 2000US-0235711P.
PR   27-SEP-2000; 2000US-0235720P.
PR   27-SEP-2000; 2000US-0235840P.
PR   27-SEP-2000; 2000US-0235863P.
PR   28-SEP-2000; 2000US-0236028P.
PR   28-SEP-2000; 2000US-0236032P.
PR   28-SEP-2000; 2000US-0236033P.
PR   28-SEP-2000; 2000US-0236034P.
PR   28-SEP-2000; 2000US-0236109P.
PR   28-SEP-2000; 2000US-0236111P.
PR   29-SEP-2000; 2000US-0236842P.
PR   29-SEP-2000; 2000US-0236891P.
PR   02-OCT-2000; 2000US-0237172P.
PR   02-OCT-2000; 2000US-0237173P.
PR   02-OCT-2000; 2000US-0237278P.
PR   02-OCT-2000; 2000US-0237294P.
PR   02-OCT-2000; 2000US-0237295P.
PR   02-OCT-2000; 2000US-0237316P.
PR   03-OCT-2000; 2000US-0237425P.
PR   03-OCT-2000; 2000US-0237598P.
PR   03-OCT-2000; 2000US-0237604P.
PR   03-OCT-2000; 2000US-0237606P.
PR   03-OCT-2000; 2000US-0237608P.
PR   01-NOV-2000; 2000US-0244867P.
PR   01-NOV-2000; 2000US-0245084P.
XX
PA   (AVAL-) AVALON PHARM.
XX
PI   Young PE,  Augustus M,  Carter KC,  Ebner R,  Endress G,  Horrigan S;
PI   Soppet DR,  Weaver Z;
XX
DR   WPI; 2002-188264/24.
XX
PT   Screening for anti-neoplastic agent involves exposing cells to a chemical
PT   agent to be tested for anti-neoplastic activity, and determining a change
PT   in expression of a gene of a signature gene set.
XX
PS   Claim 1; SEQ ID NO 3595; 44pp; English.
XX
CC   The present invention describes a method (M1) for screening for an anti-
CC   neoplastic agent. The method involves exposing cells to a chemical agent
CC   to be tested for anti-neoplastic activity, determining a change in
CC   expression of at least one gene (I) of a signature gene set, where (I)
CC   comprises a sequence (S) selected from 8447 sequences (given in ABL61664
CC   to ABL70110), or is at least 95% identical to (S), where a change in
CC   expression is indicative of anti-neoplastic activity. (I) has cytostatic
CC   activity and can be used in gene therapy. M1 can be used for screening an
CC   anti-neoplastic agent, and can be used for producing a product which is
CC   the data collected with respect to the anti-neoplastic agent as a result
CC   of M1, and the data is sufficient to convey the chemical structure and/or
CC   properties of the agent. M1 can be used in the treatment of cancer such
CC   as colon, breast, stomach, lung, thyroid, oesophageal, ovarian, kidney,
CC   prostate or pancreatic cancer, adenocarcinoma, carcinoma, clear cell
CC   cancer, infiltrating ductal cancer, infiltrating lobular cancer, squamous
CC   cell carcinoma, neuroendocrine carcinoma, papillary carcinoma and Wilm's
CC   tumour
XX
SQ   Sequence 3273 BP; 876 A; 809 C; 859 G; 729 T; 0 U; 0 Other;

  Query Match             9.9%;  Score 430.8;  DB 6;  Length 3273;
  Best Local Similarity   53.6%;  Pred. No. 3e-128;
  Matches 1114;  Conservative    0;  Mismatches  877;  Indels   87;  Gaps    7;

Qy       2181 TACATCCAGATCATAGAGGAGACTGTGGAGCGTGAGGAGGATACTCTAAATATCGACATG 2240
              ||||| ||      | | |||||  |    |      ||  |||  |     | ||||| |  ||
Db        759 TACATGCAACGACTTGGGGAGATTAATGTTATTGGTGAGCAATTTTTAAATGTGAACTGT 818

Qy       2241 TCAGACATTTATGACCATGATCCTGATCTATACGCAAAAAATTGTTCGATACCCACTCGAC 2300
                | ||||  |        |||       || | |||  | || |  |      |||||| ||
Db        819 GAACACATCAAATCATTTGACAAAAAATTTGTACAGACAACTCATCTCTTACCCACAGGAA 878

Qy       2301 ATCATCCCCCTGTTGGACACTGAGTGTCAGGAA------GTTGCTACCTCTTTACTACCA 2354
                | || ||     || ||||  |       | |||          |||     |         ||
Db        879 GTTATTCCAACTTTTGACATGGCTGTCAATGAAATCTTCTTTGACCGTTACCCTGACTCA 938

Qy       2355 ACGTTTGAGAAGCATATTGAGGCCAGACCTTTCAATCTCAAAGCATCGGTGCACATGCGT 2414
              |  || ||   | || ||| |  |   ||||| |||||          |  | |||  |
Db        939 ATCTTAGAACATCAGATTCAAGTAAGACCATTCAACGCATTGAAGACTAAGAATATGAGA 998

Qy       2415 GAACTCAACCCTTCAGATATAGACAAATTGGTTTCTGTTAAAGGAATGGTTATCCGGTGC 2474
                | || || ||      ||| || ||| |   |    |  |   || ||||| ||| ||
Db        999 AACCTGAATCCAGAAGACATTGACCAGCTCATCACCATCAGCGGCATGGTGATCAGGACA 1058

Qy       2475 AGTTCTATCATACCTGAAATTAAGGGGGCCTTCTTCAAATGTTTAGTGTGTGGTCACTCG 2534
                          | || || || ||  ||| |||||||||| | ||    ||||||||  ||| ||
Db       1059 TCCCAGCTGATTCCCGAGATGCAGGAGGCCTTCTTCCAGTGCCAAGTGTGTGCCCACACG 1118

Qy       2535 CCTCCGCTAGTTACAGTTGTTAAAGGGCGGGTTGAGGAGCCAACAAGGTGTGAAAAGCCA 2594
                | | | | |  |    | |      || ||  |||  |||||  |        ||| |
Db       1119 ACCCGGGTGGAGA---TGGACCGCGGCCGCATTGCAGAGCCCAGTGTGTGCGGGCG---- 1171

Qy       2595 GAATGTGCAGCACGGAATGCTATGTCTCTTATTCACAATCGATGCACTTTTGCAAATAAG 2654
```

```
Db   1172 --CTGCCACACCACCCACAGCATGGCACTCATCCACAACCGCTCCCTCTTCTCTGACAAG 1229

Qy   2655 CAGATAGTGCGTCTTCAAGAAACTCCAGATGCCATTCCTGAAGGAGAGACTCCACACACA 2714
Db   1230 CAGATGATCAAGCTTCAGGAGTCTCCGGAAGACATGCCTGCAGGGCAGACACCACACACA 1289

Qy   2715 GTCAGCATGTGTTTATACAACACTATGGTTGATGCTGTGAAGCCTGGAGATCGTATTGAG 2774
Db   1290 GTTATCCTGTTTGCTCACAATGATCTCGTTGACAAGGTCCAGCCTGGGGACAGAGTGAAT 1349

Qy   2775 GTAACAGGAGTTTTCAAGGCCATGGCAGTTCGAGTTGGTCCGAATCAACGAACATTACGA 2834
Db   1350 GTTACAGGCATCTATCGAGCTGTGCCTATTCGAGTCAATCCAAGAGTGAGTAATGTGAAG 1409

Qy   2835 GCATTGTATAAGACCTACATCGATTGCGTGCACGTCAAGAAGTCTGACAGGGGTCGACTG 2894
Db   1410 TCTGTCTACAAAACCCACATTGATGTCATTCATTATCGGAAAACGGATGCAAAACGTCTG 1469

Qy   2895 CAAACTGAAGATCCTATGGAGATGGATAAGGAGAATGATATGTATGCTGGGTATCATGAA 2954
Db   1470 CA--------------------------------------------TGGCCTTGATGAA 1484

Qy   2955 AGTGATACTTCAGAAGCTGCTAATGAAGCAAAGATTCAAAAACTTAAAGAGCTGTCCAAG 3014
Db   1485 GAAGCAGAACAGAAACTTTTTCAGAGAAACGTGTGGAATTGCTTAAGGAACTTTCCAGG 1544

Qy   3015 CTCCCGGGCATTTATGATAGACTTTCAAGGTCGCTGGCTCCAAGCATTTGGGAGCTTGAA 3074
Db   1545 AAACCAGACATTTATGAGAGGCTTGCTTCAGCCTTGGCTCCAAGCATTTATGAACATGAA 1604

Qy   3075 GATATTAAAAAGGGTCTTCTTTGCCAGCTCTTTGGTGGGAAGGCTAAGAAAATT------ 3128
Db   1605 GATATAAAGAAGGGAATTTTGCTTCAGCTCTTTGGCGGGACAAGGAAGGATTTTAGTCAC 1664

Qy   3129 CCATCTGGAGCATCTTTCCGAGGTGACATCAATGTTTTACTTGTTGGGGACCCTGGTACC 3188
Db   1665 ACTGGAAGGGGCAAATTTCGGGCTGAGATCAACATCTTGCTGTGTGGCGACCCTGGTACC 1724

Qy   3189 AGTAAATCTCAGCTGCTTCAGTATGTGCACAAGATAGCTCCTCGTGGAATCTACACTAGT 3248
Db   1725 AGCAAGTCCCAGCTGCTGCAGTACGTGTACAACCTCGTCCCCAGGGGCCAGTACACGTCT 1784

Qy   3249 GGGCGAGGAAGTTCGGCGGTTGGGCTGACAGCGTATGTAACGAAGGATCCAGAAACTCGA 3308
Db   1785 GGGAAGGGCTCCAGTGCAGTTGGCCTCACTGCGTACGTAATGAAAGACCCTGAGACAAGG 1844

Qy   3309 GAGACGGTATTGGAGAGCGGAGCTTTGGTTCTTAGTGATCGTGGGATATGCTGTATCGAT 3368
Db   1845 CAGCTGGTCCTGCAGACAGGTGCTCTTGTCCTGAGTGACAACGGCATCTGCTGTATCGAT 1904

Qy   3369 GAGTTCGACAAAATGTCTGATAATGCCCGAAGCATGCTTCATGAGGTAATGGAGCAACAA 3428
Db   1905 GAGTTCGACAAGATGAATGAAAGTACAAGATCGGTATTGCATGAAGTCATGGAACAGCAG 1964

Qy   3429 ACGGTATCTGTAGCCAAAGGGGGTATCATTGCCTCGCTGAACGCTCGGACGTCTGTCCTT 3488
Db   1965 ACTCTGTCCATTGCAAAGGCTGGGATCATCTGTCAGCTCAATGCGCGCACCTCTGTCCTG 2024

Qy   3489 GCATGTGCAAATCCTAGTGGGTCCCGATACAATGCGCGCCTTTCTGTGATTGATAACATC 3548
Db   2025 GCAGCAGCAAATCCCATTGAGTCTCAGTGGAATCCTAAAAAAACAACCATTGAAAACATC 2084

Qy   3549 CAGCTTCCTCCAACTCTACTTTCTAGATTTGATTTAATTTACTTAATGCTCGACAAACCA 3608
Db   2085 CAGCTGCCTCATACTTTATTATCAAGGTTTGATTTGATCTTCCTCATGCTGGACCCTCAG 2144

Qy   3609 GACGAGCAAAACGATCGTCGTCTCGCCAGGCATCTCGTGGCTTTACACTATGAAAACTAT 3668
Db   2145 GACGAAGCCTATGACAGGCGTCTGGCTCACCACCTGGTCGCACTGTACTACCAGAGCGAG 2204

Qy   3669 GAAGTTTCAAAGCAGGACGCCTTAGATCTACAAACACTTACCGCGTATATCACCTATGCT 3728
Db   2205 GAGCAGGCAGAGGAGGAGCTCCTGGACATGGCGGTGCTAAAGGACTACATTGCCTACGCG 2264

Qy   3729 CGTCAGCATGTACATCCTACATTAAGTGATGAAGCTGCTGAAGATTTGATTAATGGCTAT 3788
Db   2265 CACAGCACCATCATGCCGCGGCTAAGTGAGGAAGCCAGCCAGGCTCTCATCGAGGCTTAT 2324

Qy   3789 GTTGAGATGCGCCAAAAGGGCAACTTTCCTGGAAGCAGTAAAAAGGTGATAACAGCCACA 3848
Db   2325 GTAGACATGAGGAAGATTGGCAGTAGCCGGGGAA-----------TGGTTTCTGCATAC 2372

Qy   3849 CCTCGGCAACTCGAAAGTATGATTCGTATCAGTGAAGCCCTAGCTCGAATGAGATTTTCT 3908
Db   2373 CCTCGACAGCTAGAGTCATTAATCCGCTTAGCAGAAGCCCATGCTAAAGTAAGATTGTCT 2432

Qy   3909 GAAGTGGTAGAGAAAGTTGATGCAGCAGAAGCTGTGCGCCTTTTAGACGTCGCTTTGCAG 3968
Db   2433 AACAAAGTTGAAGCCATTGATGTGGAAGAGGCCAAACGCCTCCATCGGGAAGCTCTGAAG 2492

Qy   3969 CAATCTGCTACTGATCATGCAACAGGTACGATAGACATGGATCTTATCACGACTGGAGTG 4028
Db   2493 CAGTCTGCAACTGATCCCCGGACTGGCATCGTGGACATATCTATTCTTACTACGGGGATG 2552

Qy   4029 TCGGCCAGCGAGCGTATTCGTCGGGCCAACTTGCTAGCTGCTCTGCGAGAGCTTATAGCA 4088
Db   2553 AGTGCCACCTCTCGTAAACGGAAAGAAGAATTAGCTGAAGCATTGAAAAAGCTTATTTTA 2612

Qy   4089 GATAAAATTTCACCTGGCAGCTCCTCTGGCTTGAAGACCAGTCAGCTTCTTGAGGATATC 4148
Db   2613 TCTAA---------GGGCAAAACACCAGCTCTAAAATACCAGCAACTTTTTGAAGATATT 2663

Qy   4149 CGGAGCCAAAGCAGTGTGGACGTTAGTTTGCAGGATATTAAAAATGCTCTGGGTAGCCTC 4208
Db   2664 CGGGGACAATCTGACATAGCAATTACTAAAGATATGTTTGAAGAAGCACTGCGTGCCCTG 2723

Qy   4209 CAAGGAGAAGGCTTTCTTACTGTCCATGGTGACATAGT 4246
Db   2724 GCAGATGATGATTTCCTGACAGTGACTGGGAAGACCGT 2761
```

```
RESULT 14
ABL65666
ID   ABL65666 standard; DNA; 3273 BP.
XX
AC   ABL65666;
XX
DT   15-MAY-2002  (first entry)
XX
DE   Lung cancer related gene sequence SEQ ID NO:4003.
XX
KW   Human; cancer; colon; breast; ovary; oesophagus; kidney; thyroid;
KW   stomach; lung; prostate; pancreas; carcinoma; antitumour; cancerous;
KW   cytostatic; gene therapy; antineoplastic; Wilm's tumour; adenocarcinoma;
KW   gene; ds.
XX
OS   Homo sapiens.
XX
PN   WO200194629-A2.
XX
PD   13-DEC-2001.
XX
PF   30-MAY-2001; 2001WO-US010838.
XX
PR   05-JUN-2000; 2000US-0209473P.
PR   05-JUN-2000; 2000US-0209531P.
PR   18-SEP-2000; 2000US-0233133P.
PR   18-SEP-2000; 2000US-0233617P.
PR   20-SEP-2000; 2000US-0234009P.
PR   20-SEP-2000; 2000US-0234034P.
PR   20-SEP-2000; 2000US-0234052P.
PR   22-SEP-2000; 2000US-0234509P.
PR   22-SEP-2000; 2000US-0234567P.
PR   25-SEP-2000; 2000US-0234923P.
PR   25-SEP-2000; 2000US-0234924P.
```

```
PR   25-SEP-2000; 2000US-0235077P.
PR   25-SEP-2000; 2000US-0235082P.
PR   25-SEP-2000; 2000US-0235134P.
PR   25-SEP-2000; 2000US-0235280P.
PR   26-SEP-2000; 2000US-0235637P.
PR   26-SEP-2000; 2000US-0235638P.
PR   27-SEP-2000; 2000US-0235711P.
PR   27-SEP-2000; 2000US-0235720P.
PR   27-SEP-2000; 2000US-0235840P.
PR   27-SEP-2000; 2000US-0235863P.
PR   28-SEP-2000; 2000US-0236028P.
PR   28-SEP-2000; 2000US-0236032P.
PR   28-SEP-2000; 2000US-0236033P.
PR   28-SEP-2000; 2000US-0236034P.
PR   28-SEP-2000; 2000US-0236109P.
PR   28-SEP-2000; 2000US-0236111P.
PR   29-SEP-2000; 2000US-0236842P.
PR   29-SEP-2000; 2000US-0236891P.
PR   02-OCT-2000; 2000US-0237172P.
PR   02-OCT-2000; 2000US-0237173P.
PR   02-OCT-2000; 2000US-0237278P.
PR   02-OCT-2000; 2000US-0237294P.
PR   02-OCT-2000; 2000US-0237295P.
PR   02-OCT-2000; 2000US-0237316P.
PR   03-OCT-2000; 2000US-0237425P.
PR   03-OCT-2000; 2000US-0237598P.
PR   03-OCT-2000; 2000US-0237604P.
PR   03-OCT-2000; 2000US-0237606P.
PR   03-OCT-2000; 2000US-0237608P.
PR   01-NOV-2000; 2000US-0244867P.
PR   01-NOV-2000; 2000US-0245084P.
XX
PA   (AVAL-) AVALON PHARM.
XX
PI   Young PE,  Augustus M,  Carter KC,  Ebner R,  Endress G,  Horrigan S;
PI   Soppet DR,  Weaver Z;
XX
DR   WPI; 2002-188264/24.
XX
PT   Screening for anti-neoplastic agent involves exposing cells to a chemical
PT   agent to be tested for anti-neoplastic activity, and determining a change
PT   in expression of a gene of a signature gene set.
XX
PS   Claim 1; SEQ ID NO 4003; 44pp; English.
XX
CC   The present invention describes a method (M1) for screening for an anti-
CC   neoplastic agent. The method involves exposing cells to a chemical agent
CC   to be tested for anti-neoplastic activity, determining a change in
CC   expression of at least one gene (I) of a signature gene set, where (I)
CC   comprises a sequence (S) selected from 8447 sequences (given in ABL61664
CC   to ABL70110), or is at least 95% identical to (S), where a change in
CC   expression is indicative of anti-neoplastic activity. (I) has cytostatic
CC   activity and can be used in gene therapy. M1 can be used for screening an
CC   anti-neoplastic agent, and can be used for producing a product which is
CC   the data collected with respect to the anti-neoplastic agent as a result
CC   of M1, and the data is sufficient to convey the chemical structure and/or
CC   properties of the agent. M1 can be used in the treatment of cancer such
CC   as colon, breast, stomach, lung, thyroid, oesophageal, ovarian, kidney,
CC   prostate or pancreatic cancer, adenocarcinoma, carcinoma, clear cell
CC   cancer, infiltrating ductal cancer, infiltrating lobular cancer, squamous
CC   cell carcinoma, neuroendocrine carcinoma, papillary carcinoma and Wilm's
CC   tumour
XX
SQ   Sequence 3273 BP; 876 A; 809 C; 859 G; 729 T; 0 U; 0 Other;

  Query Match             9.9%;  Score 430.8;  DB 6;  Length 3273;
  Best Local Similarity   53.6%;  Pred. No. 3e-128;
  Matches 1114;  Conservative    0;  Mismatches  877;  Indels   87;  Gaps    7;

Qy       2181 TACATCCAGATCATAGAGGAGACTGTGGAGCGTGAGGAGGATACTCTAAATATCGACATG 2240
Db        759 TACATGCAACGACTTGGGGAGATTAATGTTATTGGTGAGCAATTTTTAAATGTGAACTGT 818

Qy       2241 TCAGACATTTATGACCATGATCCTGATCTATACGCAAAAATTGTTCGATACCCACTCGAC 2300
Db        819 GAACACATCAAATCATTTGACAAAAAATTTGTACAGACAACTCATCTCTTACCCACAGGAA 878

Qy       2301 ATCATCCCCCTGTTGGACACTGAGTGTCAGGAA------GTTGCTACCTCTTTACTACCA 2354
Db        879 GTTATTCCAACTTTTGACATGGCTGTCAATGAAATCTTCTTTGACCGTTACCCTGACTCA 938

Qy       2355 ACGTTTGAGAAGCATATTGAGGCCAGACCTTTCAATCTCAAAGCATCGGTGCACATGCGT 2414
Db        939 ATCTTAGAACATCAGATTCAAGTAAGACCATTCAACGCATTGAAGACTAAGAATATGAGA 998

Qy       2415 GAACTCAACCCTTCAGATATAGACAAATTGGTTTCTGTTAAAGGAATGGTTATCCGGTGC 2474
Db        999 AACCTGAATCCAGAAGACATTGACCAGCTCATCACCATCAGCGGCATGGTGATCAGGACA 1058

Qy       2475 AGTTCTATCATACCTGAAATTAAGGGGGCCTTCTTCAAATGTTTAGTGTGTGGTCACTCG 2534
Db       1059 TCCCAGCTGATTCCCGAGATGCAGGAGGCCTTCTTCCAGTGCCAAGTGTGTGCCCACACG 1118

Qy       2535 CCTCCGCTAGTTACAGTTGTTAAAGGGCGGGTTGAGGAGCCAACAAGGTGTGAAAAGCCA 2594
Db       1119 ACCCGGGTGGAGA---TGGACCGCGGCCGCATTGCAGAGCCCAGTGTGTGCGGGCG---- 1171

Qy       2595 GAATGTGCAGCACGGAATGCTATGTCTCTTATTCACAATCGATGCACTTTTGCAAATAAG 2654
Db       1172 --CTGCCACACCACCCACAGCATGGCACTCATCCACAACCGCTCCCTCTTCTCTGACAAG 1229

Qy       2655 CAGATAGTGCGTCTTCAAGAAACTCCAGATGCCATTCCTGAAGGAGAGACTCCACACACA 2714
Db       1230 CAGATGATCAAGCTTCAGGAGTCTCCGGAAGACATGCCTGCAGGGCAGACACCACACACA 1289

Qy       2715 GTCAGCATGTGTTTATACAACACTATGGTTGATGCTGTGAAGCCTGGAGATCGTATTGAG 2774
Db       1290 GTTATCCTGTTTGCTCACAATGATCTCGTTGACAAGGTCCAGCCTGGGGACAGAGTGAAT 1349

Qy       2775 GTAACAGGAGTTTTCAAGGCCATGGCAGTTCGAGTTGGTCCGAATCAACGAACATTACGA 2834
Db       1350 GTTACAGGCATCTATCGAGCTGTGCCTATTCGAGTCAATCCAAGAGTGAGTAATGTGAAG 1409

Qy       2835 GCATTGTATAAGACCTACATCGATTGCGTGCACGTCAAGAAGTCTGACAGGGGTCGACTG 2894
Db       1410 TCTGTCTACAAAACCCACATTGATGTCATTCATTATCGGAAAACGGATGCAAAACGTCTG 1469

Qy       2895 CAAACTGAAGATCCTATGGAGATGGATAAGGAGAATGATATGTATGCTGGGTATCATGAA 2954
Db       1470 CA--------------------------------------------TGGCCTTGATGAA 1484

Qy       2955 AGTGATACTTCAGAAGCTGCTAATGAAGCAAAGATTCAAAAACTTAAAGAGCTGTCCAAG 3014
Db       1485 GAAGCAGAACAGAAACTTTTTTCAGAGAAACGTGTGGAATTGCTTAAGGAACTTTCCAGG 1544

Qy       3015 CTCCCGGGCATTTATGATAGACTTTCAAGGTCGCTGGCTCCAAGCATTTGGGAGCTTGAA 3074
Db       1545 AAACCAGACATTTATGAGAGGCTTGCTTCAGCCTTGGCTCCAAGCATTTATGAACATGAA 1604

Qy       3075 GATATTAAAAAGGGTCTTCTTTGCCAGCTCTTTGGTGGAAGGCTAAGAAAATT------ 3128
Db       1605 GATATAAAGAAGGGAATTTTGCTTCAGCTCTTTGGCGGGACAAGGAAGGATTTTAGTCAC 1664

Qy       3129 CCATCTGGAGCATCTTTCCGAGGTGACATCAATGTTTTACTTGTTGGGGACCCTGGTACC 3188
Db       1665 ACTGGAAGGGGCAAATTTCGGGCTGAGATCAACATCTTGCTGTGTGGCGACCCTGGTACC 1724

Qy       3189 AGTAAATCTCAGCTGCTTCAGTATGTGCACAAGATAGCTCCTCGTGGAATCTACACTAGT 3248
Db       1725 AGCAAGTCCCAGCTGCTGCAGTACGTGTACAACCTCGTCCCCAGGGGCCAGTACACGTCT 1784

Qy       3249 GGGCGAGGAAGTTCGGCGGTTGGGCTGACAGCGTATGTAACGAAGGATCCAGAAACTCGA 3308
Db       1785 GGGAAGGGCTCCAGTGCAGTTGGCCTCACTGCGTACGTAATGAAAGACCCTGAGACAAGG 1844

Qy       3309 GAGACGGTATTGGAGAGCGGAGCTTTGGTTCTTAGTGATCGTGGGATATGCTGTATCGAT 3368
```

```
Db       1845 CAGCTGGTCCTGCAGACAGGTGCTCTTGTCCTGAGTGACAACGGCATCTGCTGTATCGAT 1904

Qy       3369 GAGTTCGACAAAATGTCTGATAATGCCCGAAGCATGCTTCATGAGGTAATGGAGCAACAA 3428

Db       1905 GAGTTCGACAAGATGAATGAAAGTACAAGATCGGTATTGCATGAAGTCATGGAACAGCAG 1964

Qy       3429 ACGGTATCTGTAGCCAAAGGGGGTATCATTGCCTCGCTGAACGCTCGGACGTCTGTCCTT 3488

Db       1965 ACTCTGTCCATTGCAAAGGCTGGGATCATCTGTCAGCTCAATGCGCGCACCTCTGTCCTG 2024

Qy       3489 GCATGTGCAAATCCTAGTGGGTCCCGATACAATGCGCGCCTTTCTGTGATTGATAACATC 3548

Db       2025 GCAGCAGCAAATCCCATTGAGTCTCAGTGGAATCCTAAAAAAACAACCATTGAAAACATC 2084

Qy       3549 CAGCTTCCTCCAACTCTACTTTCTAGATTTGATTTAATTTACTTAATGCTCGACAAACCA 3608

Db       2085 CAGCTGCCTCATACTTTATTATCAAGGTTTGATTTGATCTTCCTCATGCTGGACCCTCAG 2144

Qy       3609 GACGAGCAAAACGATCGTCGTCTCGCCAGGCATCTCGTGGCTTTACACTATGAAAACTAT 3668

Db       2145 GACGAAGCCTATGACAGGCGTCTGGCTCACCACCTGGTCGCACTGTACTACCAGAGCGAG 2204

Qy       3669 GAAGTTTCAAAGCAGGACGCCTTAGATCTACAAACACTTACCGCGTATATCACCTATGCT 3728

Db       2205 GAGCAGGCAGAGGAGGAGCTCCTGGACATGGCGGTGCTAAAGGACTACATTGCCTACGCG 2264

Qy       3729 CGTCAGCATGTACATCCTACATTAAGTGATGAAGCTGCTGAAGATTTGATTAATGGCTAT 3788

Db       2265 CACAGCACCATCATGCCGCGGCTAAGTGAGGAAGCCAGCCAGGCTCTCATCGAGGCTTAT 2324

Qy       3789 GTTGAGATGCGCCAAAAGGGCAACTTTCCTGGAAGCAGTAAAAAGGTGATAACAGCCACA 3848

Db       2325 GTAGACATGAGGAAGATTGGCAGTAGCCGGGGAA------------TGGTTTCTGCATAC 2372

Qy       3849 CCTCGGCAACTCGAAAGTATGATTCGTATCAGTGAAGCCCTAGCTCGAATGAGATTTTCT 3908

Db       2373 CCTCGACAGCTAGAGTCATTAATCCGCTTAGCAGAAGCCCATGCTAAAGTAAGATTGTCT 2432

Qy       3909 GAAGTGGTAGAGAAAGTTGATGCAGCAGAAGCTGTGCGCCTTTTAGACGTCGCTTTGCAG 3968

Db       2433 AACAAAGTTGAAGCCATTGATGTGGAAGAGGCCAAACGCCTCCATCGGGAAGCTCTGAAG 2492

Qy       3969 CAATCTGCTACTGATCATGCAACAGGTACGATAGACATGGATCTTATCACGACTGGAGTG 4028

Db       2493 CAGTCTGCAACTGATCCCCGGACTGGCATCGTGGACATATCTATTCTTACTACGGGGATG 2552

Qy       4029 TCGGCCAGCGAGCGTATTCGTCGGCCAACTTGCTAGCTGCTCTGCGAGAGCTTATAGCA 4088

Db       2553 AGTGCCACCTCTCGTAAACGGAAAGAAGAATTAGCTGAAGCATTGAAAAAGCTTATTTA 2612

Qy       4089 GATAAAATTTCACCTGGCAGCTCCTCTGGCTTGAAGACCAGTCAGCTTCTTGAGGATATC 4148

Db       2613 TCTAA---------GGGCAAAACACCAGCTCTAAAATACCAGCAACTTTTTGAAGATATT 2663

Qy       4149 CGGAGCCAAAGCAGTGTGGACGTTAGTTTGCAGGATATTAAAAATGCTCTGGGTAGCCTC 4208

Db       2664 CGGGGACAATCTGACATAGCAATTACTAAAGATATGTTTGAAGAAGCACTGCGTGCCCTG 2723

Qy       4209 CAAGGAGAAGGCTTTCTTACTGTCCATGGTGACATAGT 4246

Db       2724 GCAGATGATGATTTCCTGACAGTGACTGGGAAGACCGT 2761
```

RESULT 15
ABL65259
ID   ABL65259 standard; DNA; 3273 BP.
XX
AC   ABL65259;
XX
DT   15-MAY-2002 (first entry)
XX
DE   Lung cancer related gene sequence SEQ ID NO:3596.
XX
KW   Human; cancer; colon; breast; ovary; oesophagus; kidney; thyroid;
KW   stomach; lung; prostate; pancreas; carcinoma; antitumour; cancerous;
KW   cytostatic; gene therapy; antineoplastic; Wilm's tumour; adenocarcinoma;
KW   gene; ds.
XX
OS   Homo sapiens.
XX
PN   WO200194629-A2.
XX
PD   13-DEC-2001.
XX
PF   30-MAY-2001; 2001WO-US010838.
XX
PR   05-JUN-2000; 2000US-0209473P.
PR   05-JUN-2000; 2000US-0209531P.
PR   18-SEP-2000; 2000US-0233133P.
PR   18-SEP-2000; 2000US-0233617P.
PR   20-SEP-2000; 2000US-0234009P.
PR   20-SEP-2000; 2000US-0234034P.
PR   20-SEP-2000; 2000US-0234052P.
PR   22-SEP-2000; 2000US-0234509P.
PR   22-SEP-2000; 2000US-0234567P.
PR   25-SEP-2000; 2000US-0234923P.
PR   25-SEP-2000; 2000US-0234924P.
PR   25-SEP-2000; 2000US-0235077P.
PR   25-SEP-2000; 2000US-0235082P.
PR   25-SEP-2000; 2000US-0235134P.
PR   25-SEP-2000; 2000US-0235280P.
PR   26-SEP-2000; 2000US-0235637P.
PR   26-SEP-2000; 2000US-0235638P.
PR   27-SEP-2000; 2000US-0235711P.
PR   27-SEP-2000; 2000US-0235720P.
PR   27-SEP-2000; 2000US-0235840P.
PR   27-SEP-2000; 2000US-0235863P.
PR   28-SEP-2000; 2000US-0236028P.
PR   28-SEP-2000; 2000US-0236032P.
PR   28-SEP-2000; 2000US-0236033P.
PR   28-SEP-2000; 2000US-0236034P.
PR   28-SEP-2000; 2000US-0236109P.
PR   28-SEP-2000; 2000US-0236111P.
PR   29-SEP-2000; 2000US-0236842P.
PR   29-SEP-2000; 2000US-0236891P.
PR   02-OCT-2000; 2000US-0237172P.
PR   02-OCT-2000; 2000US-0237173P.
PR   02-OCT-2000; 2000US-0237278P.
PR   02-OCT-2000; 2000US-0237294P.
PR   02-OCT-2000; 2000US-0237295P.
PR   02-OCT-2000; 2000US-0237316P.
PR   03-OCT-2000; 2000US-0237425P.
PR   03-OCT-2000; 2000US-0237598P.
PR   03-OCT-2000; 2000US-0237604P.
PR   03-OCT-2000; 2000US-0237606P.
PR   03-OCT-2000; 2000US-0237608P.
PR   01-NOV-2000; 2000US-0244867P.
PR   01-NOV-2000; 2000US-0245084P.
XX
PA   (AVAL-) AVALON PHARM.
XX
PI   Young PE, Augustus M, Carter KC, Ebner R, Endress G, Horrigan S;
PI   Soppet DR, Weaver Z;
XX
DR   WPI; 2002-188264/24.
XX
PT   Screening for anti-neoplastic agent involves exposing cells to a chemical
PT   agent to be tested for anti-neoplastic activity, and determining a change
PT   in expression of a gene of a signature gene set.
XX
PS   Claim 1; SEQ ID NO 3596; 44pp; English.
XX
CC   The present invention describes a method (M1) for screening for an anti-
CC   neoplastic agent. The method involves exposing cells to a chemical agent
CC   to be tested for anti-neoplastic activity, determining a change in

```
CC   expression of at least one gene (I) of a signature gene set, where (I)
CC   comprises a sequence (S) selected from 8447 sequences (given in ABL61664
CC   to ABL70110), or is at least 95% identical to (S), where a change in
CC   expression is indicative of anti-neoplastic activity. (I) has cytostatic
CC   activity and can be used in gene therapy. M1 can be used for screening an
CC   anti-neoplastic agent, and can be used for producing a product which is
CC   the data collected with respect to the anti-neoplastic agent as a result
CC   of M1, and the data is sufficient to convey the chemical structure and/or
CC   properties of the agent. M1 can be used in the treatment of cancer such
CC   as colon, breast, stomach, lung, thyroid, oesophageal, ovarian, kidney,
CC   prostate or pancreatic cancer, adenocarcinoma, carcinoma, clear cell
CC   cancer, infiltrating ductal cancer, infiltrating lobular cancer, squamous
CC   cell carcinoma, neuroendocrine carcinoma, papillary carcinoma and Wilm's
CC   tumour
XX
SQ   Sequence 3273 BP; 876 A; 809 C; 859 G; 729 T; 0 U; 0 Other;

  Query Match             9.9%;  Score 430.8;  DB 6;  Length 3273;
  Best Local Similarity   53.6%;  Pred. No. 3e-128;
  Matches 1114;  Conservative    0;  Mismatches  877;  Indels   87;  Gaps    7;

Qy       2181 TACATCCAGATCATAGAGGAGACTGTGGAGCGTGAGGAGGATACTCTAAATATCGACATG 2240
Db        759 TACATGCAACGACTTGGGGAGATTAATGTTATTGGTGAGCAATTTTTAAATGTGAACTGT 818
Qy       2241 TCAGACATTTATGACCATGATCCTGATCTATACGCAAAAATTGTTCGATACCCACTCGAC 2300
Db        819 GAACACATCAAATCATTTGACAAAAATTTGTACAGACAACTCATCTCTTACCCACAGGAA 878
Qy       2301 ATCATCCCCCTGTTGGACACTGAGTGTCAGGAA------GTTGCTACCTCTTTACTACCA 2354
Db        879 GTTATTCCAACTTTTGACATGGCTGTCAATGAAATCTTCTTTGACCGTTACCCTGACTCA 938
Qy       2355 ACGTTTGAGAAGCATATTGAGGCCAGACCTTTCAATCTCAAAGCATCGGTGCACATGCGT 2414
Db        939 ATCTTAGAACATCAGATTCAAGTAAGACCATTCAACGCATTGAAGACTAAGAATATGAGA 998
Qy       2415 GAACTCAACCCTTCAGATATAGACAAATTGGTTTCTGTTAAAGGAATGGTTATCCGGTGC 2474
Db        999 AACCTGAATCCAGAAGACATTGACCAGCTCATCACCATCAGCGGCATGGTGATCAGGACA 1058
Qy       2475 AGTTCTATCATACCTGAAATTAAGGGGGCCTTCTTCAAATGTTTAGTGTGTGGTCACTCG 2534
Db       1059 TCCCAGCTGATTCCCGAGATGCAGGAGGCCTTCTTCCAGTGCCAAGTGTGTGCCCACACG 1118
Qy       2535 CCTCCGCTAGTTACAGTTGTTAAAGGGCGGGTTGAGGAGCCAACAAGGTGTGAAAAGCCA 2594
Db       1119 ACCCGGGTGGAGA---TGGACCGCGGCCGCATTGCAGAGCCCAGTGTGTGCGGGCG---- 1171
Qy       2595 GAATGTGCAGCACGGAATGCTATGTCTCTTATTCACAATCGATGCACTTTTGCAAATAAG 2654
Db       1172 --CTGCCACACCACCCACAGCATGGCACTCATCCACAACCGCTCCCTCTTCTCTGACAAG 1229
Qy       2655 CAGATAGTGCGTCTTCAAGAAACTCCAGATGCCATTCCTGAAGGAGAGACTCCACACACA 2714
Db       1230 CAGATGATCAAGCTTCAGGAGTCTCCGGAAGACATGCCTGCAGGGCAGACACCACACACA 1289
Qy       2715 GTCAGCATGTGTTTATACAACACTATGGTTGATGCTGTGAAGCCTGGAGATCGTATTGAG 2774
Db       1290 GTTATCCTGTTTGCTCACAATGATCTCGTTGACAAGGTCCAGCCTGGGGACAGAGTGAAT 1349
Qy       2775 GTAACAGGAGTTTTCAAGGCCATGGCAGTTCGAGTTGGTCCGAATCAACGAACATTACGA 2834
Db       1350 GTTACAGGCATCTATCGAGCTGTGCCTATTCGAGTCAATCCAAGAGTGAGTAATGTGAAG 1409
Qy       2835 GCATTGTATAAGACCTACATCGATTGCGTGCACGTCAAGAAGTCTGACAGGGGTCGACTG 2894
Db       1410 TCTGTCTACAAAACCCACATTGATGTCATTCATTATCGGAAAACGGATGCAAAACGTCTG 1469
Qy       2895 CAAACTGAAGATCCTATGGAGATGGATAAGGAGAATGATATGTATGCTGGGTATCATGAA 2954
Db       1470 CA------------------------------------------TGGCCTTGATGAA 1484

Qy       2955 AGTGATACTTCAGAAGCTGCTAATGAAGCAAAGATTCAAAAACTTAAAGAGCTGTCCAAG 3014
Db       1485 GAAGCAGAACAGAAACTTTTTCAGAGAAACGTGTGGAATTGCTTAAGGAACTTTCCAGG 1544
Qy       3015 CTCCCGGGCATTTATGATAGACTTTCAAGGTCGCTGGCTCCAAGCATTTGGGAGCTTGAA 3074
Db       1545 AAACCAGACATTTATGAGAGGCTTGCTTCAGCCTTGGCTCCAAGCATTTATGAACATGAA 1604
Qy       3075 GATATTAAAAAGGGTCTTCTTTGCCAGCTCTTTGGTGGGAAGGCTAAGAAAATT------ 3128
Db       1605 GATATAAAGAAGGGAATTTTGCTTCAGCTCTTTGGCGGGACAAGGAAGGATTTTAGTCAC 1664
Qy       3129 CCATCTGGAGCATCTTTCCGAGGTGACATCAATGTTTTACTTGTTGGGGACCCTGGTACC 3188
Db       1665 ACTGGAAGGGGCAAATTTCGGGCTGAGATCAACATCTTGCTGTGTGGCGACCCTGGTACC 1724
Qy       3189 AGTAAATCTCAGCTGCTTCAGTATGTGCACAAGATAGCTCCTCGTGGAATCTACACTAGT 3248
Db       1725 AGCAAGTCCCAGCTGCTGCAGTACGTGTACAACCTCGTCCCCAGGGGCCAGTACACGTCT 1784
Qy       3249 GGGCGAGGAAGTTCGGCGGTTGGGCTGACAGCGTATGTAACGAAGGATCCAGAAACTCGA 3308
Db       1785 GGGAAGGGCTCCAGTGCAGTTGGCCTCACTGCGTACGTAATGAAAGACCCTGAGACAAGG 1844
Qy       3309 GAGACGGTATTGGAGAGCGGAGCTTTGGTTCTTAGTGATCGTGGGATATGCTGTATCGAT 3368
Db       1845 CAGCTGGTCCTGCAGACAGGTGCTCTTGTCCTGAGTGACAACGGCATCTGCTGTATCGAT 1904
Qy       3369 GAGTTCGACAAAATGTCTGATAATGCCCGAAGCATGCTTCATGAGGTAATGGAGCAACAA 3428
Db       1905 GAGTTCGACAAGATGAATGAAAGTACAAGATCGGTATTGCATGAAGTCATGGAACAGCAG 1964
Qy       3429 ACGGTATCTGTAGCCAAAGGGGGTATCATTGCCTCGCTGAACGCTCGGACGTCTGTCCTT 3488
Db       1965 ACTCTGTCCATTGCAAAGGCTGGGATCATCTGTCAGCTCAATGCGCGCACCTCTGTCCTG 2024
Qy       3489 GCATGTGCAAATCCTAGTGGGTCCCGATACAATGCGCGCCTTTCTGTGATTGATAACATC 3548
Db       2025 GCAGCAGCAAATCCCATTGAGTCTCAGTGGAATCCTAAAAAAACAACCATTGAAAACATC 2084
Qy       3549 CAGCTTCCTCCAACTCTACTTTCTAGATTTGATTTAATTTACTTAATGCTCGACAAACCA 3608
Db       2085 CAGCTGCCTCATACTTTATTATCAAGGTTTGATTTGATCTTCCTCATGCTGGACCCTCAG 2144
Qy       3609 GACGAGCAAAACGATCGTCGTCTCGCCAGGCATCTCGTGGCTTTACACTATGAAAACTAT 3668
Db       2145 GACGAAGCCTATGACAGGCGTCTGGCTCACCACCTGGTCGCACTGTACTACCAGAGCGAG 2204
Qy       3669 GAAGTTTCAAAGCAGGACGCCTTAGATCTACAAACACTTACCGCGTATATCACCTATGCT 3728
Db       2205 GAGCAGGCAGAGGAGGAGCTCCTGGACATGGCGGTGCTAAAGGACTACATTGCCTACGCG 2264
Qy       3729 CGTCAGCATGTACATCCTACATTAAGTGATGAAGCTGCTGAAGATTTGATTAATGGCTAT 3788
Db       2265 CACAGCACCATCATGCCGCGGCTAAGTGAGGAAGCCAGCCAGGCTCTCATCGAGGCTTAT 2324
Qy       3789 GTTGAGATGCGCCAAAAGGGCAACTTTCCTGGAAGCAGTAAAAAGGTGATAACAGCCACA 3848
Db       2325 GTAGACATGAGGAAGATTGGCAGTAGCCGGGGAA-----------TGGTTTCTGCATAC 2372
Qy       3849 CCTCGGCAACTCGAAAGTATGATTCGTATCAGTGAAGCCCTAGCTCGAATGAGATTTTCT 3908
Db       2373 CCTCGACAGCTAGAGTCATTAATCCGCTTAGCAGAAGCCCATGCTAAAGTAAGATTGTCT 2432
Qy       3909 GAAGTGGTAGAGAAAGTTGATGCAGCAGAAGCTGTGCGCCTTTTAGACGTCGCTTTGCAG 3968
Db       2433 AACAAAGTTGAAGCCATTGATGTGGAAGAGGCCAAACGCCTCCATCGGGAAGCTCTGAAG 2492
Qy       3969 CAATCTGCTACTGATCATGCAACAGGTACGATAGACATGGATCTTATCACGACTGGAGTG 4028
Db       2493 CAGTCTGCAACTGATCCCCGGACTGGCATCGTGGACATATCTATTCTTACTACGGGGATG 2552
Qy       4029 TCGGCCAGCGAGCGTATTCGTCGGGCCAACTTGCTAGCTGCTCTGCGAGAGCTTATAGCA 4088
```

```
                 |||| |   |||| ||    |   | ||     |  ||  ||  | |||||||    |
Db       2553 AGTGCCACCTCTCGTAAACGGAAAGAAGAATTAGCTGAAGCATTGAAAAAGCTTATTTTA 2612

Qy       4089 GATAAAATTTCACCTGGCAGCTCCTCTGGCTTGAAGACCAGTCAGCTTCTTGAGGATATC 4148
                 |||           ||||   |  |  |   | ||    |    || ||| |||| |||||
Db       2613 TCTAA---------GGGCAAAACACCAGCTCTAAAATACCAGCAACTTTTTGAAGATATT 2663

Qy       4149 CGGAGCCAAAGCAGTGTGGACGTTAGTTTGCAGGATATTAAAAATGCTCTGGGTAGCCTC 4208
              ||| | |||        | |    ||| |    |       || || | || ||| ||  |||
Db       2664 CGGGGACAATCTGACATAGCAATTACTAAAGATATGTTTGAAGAAGCACTGCGTGCCCTG 2723

Qy       4209 CAAGGAGAAGGCTTTCTTACTGTCCATGGTGACATAGT 4246
                 ||  || |  || || || ||   |||  | |  ||
Db       2724 GCAGATGATGATTTCCTGACAGTGACTGGGAAGACCGT 2761
```

Search completed: December 6, 2005, 02:17:03
Job time : 1616 secs

```
                    GenCore version 5.1.6
          Copyright (c) 1993 - 2005  Compugen Ltd.

OM nucleic - nucleic search, using sw model

Run on:         December  6, 2005, 01:47:40 ; Search time 504 Seconds
                                          (without alignments)
                                          15334.996 Million cell updates/sec

Title:          US-10-768-511-5
Perfect score:  4348
Sequence:       1 atggcgcgccgcactcacgt..........gcgcaattcccagagctcgc 4348

Scoring table:  IDENTITY_NUC
                Gapop 10.0 , Gapext 1.0

Searched:       1303057 seqs, 888780828 residues

Total number of hits satisfying chosen parameters:      2606114

Minimum DB seq length: 0
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%
                 Maximum Match 100%
                 Listing first 45 summaries

Database :       Issued_Patents_NA:*
                1:  /cgn2_6/ptodata/1/ina/1_COMB.seq:*
                2:  /cgn2_6/ptodata/1/ina/5_COMB.seq:*
                3:  /cgn2_6/ptodata/1/ina/6A_COMB.seq:*
                4:  /cgn2_6/ptodata/1/ina/6B_COMB.seq:*
                5:  /cgn2_6/ptodata/1/ina/H_COMB.seq:*
                6:  /cgn2_6/ptodata/1/ina/PCTUS_COMB.seq:*
                7:  /cgn2_6/ptodata/1/ina/PP_COMB.seq:*
                8:  /cgn2_6/ptodata/1/ina/RE_COMB.seq:*
                9:  /cgn2_6/ptodata/1/ina/backfiles1.seq:*

     Pred. No. is the number of results predicted by chance to have a
     score greater than or equal to the score of the result being printed,
     and is derived by analysis of the total score distribution.

                              SUMMARIES
```

	Result No.	Score	% Query Match	Length	DB	ID	Description
	1	4348	100.0	4348	3	US-09-828-062-5	Sequence 5, Appli
	2	673	15.5	723	3	US-09-828-062-2	Sequence 2, Appli
	3	209.6	4.8	2533	3	US-09-604-605-1	Sequence 1, Appli
	4	209.6	4.8	2533	3	US-10-101-080-1	Sequence 1, Appli
	5	198.6	4.6	1635	3	US-09-248-796A-4534	Sequence 4534, Ap
	6	197.6	4.5	1488	3	US-09-248-796A-5904	Sequence 5904, Ap
	7	186.4	4.3	2917	3	US-09-949-016-2468	Sequence 2468, Ap
	8	183.8	4.2	2533	3	US-09-949-016-5021	Sequence 5021, Ap
	9	180	4.1	3000	3	US-09-949-016-2621	Sequence 2621, Ap
	10	173.2	4.0	996	3	US-09-248-796A-4570	Sequence 4570, Ap
	11	160.8	3.7	3378	3	US-09-964-899-48	Sequence 48, Appl
	12	160.8	3.7	3445	3	US-09-976-594-323	Sequence 323, App
	13	157.6	3.6	3379	3	US-09-220-132-12	Sequence 12, Appl
	14	137	3.2	294	3	US-09-313-294A-4610	Sequence 4610, Ap
	15	135.6	3.1	311	3	US-09-313-294A-83	Sequence 83, Appl
	16	133.8	3.1	297	3	US-09-313-294A-4978	Sequence 4978, Ap
	17	127.8	2.9	12932	3	US-09-949-016-14363	Sequence 14363, A
	18	123.2	2.8	507	3	US-09-401-064-158	Sequence 158, App
	19	120.2	2.8	519	3	US-09-248-796A-5916	Sequence 5916, Ap
	20	105.2	2.4	1305	3	US-09-248-796A-4711	Sequence 4711, Ap
	21	96.4	2.2	256	3	US-09-313-294A-2519	Sequence 2519, Ap
	22	78.4	1.8	492	3	US-09-248-796A-5915	Sequence 5915, Ap
	23	73	1.7	273	3	US-09-313-294A-1684	Sequence 1684, Ap
c	24	70.6	1.6	1664976	3	US-08-916-421B-1	Sequence 1, Appli
c	25	70.6	1.6	1664976	3	US-09-692-570-1	Sequence 1, Appli
	26	69.4	1.6	613	3	US-09-533-559-658	Sequence 658, App
	27	59.6	1.4	28366	3	US-09-949-016-16763	Sequence 16763, A
	28	59.4	1.4	39949	3	US-09-949-016-14210	Sequence 14210, A
c	29	53.6	1.2	14636	3	US-09-173-914-6	Sequence 6, Appli
	30	51	1.2	1290	3	US-09-248-796A-4988	Sequence 4988, Ap
	31	49	1.1	2313	3	US-09-799-451-937	Sequence 937, App
	32	46.8	1.1	289	3	US-09-313-294A-3770	Sequence 3770, Ap
c	33	46.4	1.1	565	3	US-09-328-111-454	Sequence 454, App
	34	45	1.0	601	3	US-09-949-016-176386	Sequence 176386,
c	35	44.4	1.0	7218	2	US-08-232-463-14	Sequence 14, Appl
	36	41.2	0.9	1572	3	US-09-328-352-2268	Sequence 2268, Ap
	37	40.2	0.9	818128	3	US-09-949-016-14546	Sequence 14546, A
	38	40.2	0.9	818128	3	US-09-949-016-14547	Sequence 14547, A
	39	40.2	0.9	818128	3	US-09-949-016-14548	Sequence 14548, A
	40	40.2	0.9	818128	3	US-09-949-016-14549	Sequence 14549, A
	41	40.2	0.9	818128	3	US-09-949-016-14550	Sequence 14550, A
	42	40.2	0.9	818128	3	US-09-949-016-14551	Sequence 14551, A
	43	40.2	0.9	818128	3	US-09-949-016-14552	Sequence 14552, A
	44	40.2	0.9	818128	3	US-09-949-016-14553	Sequence 14553, A
	45	40.2	0.9	818128	3	US-09-949-016-14554	Sequence 14554, A

```
                              ALIGNMENTS

RESULT 1
US-09-828-062-5
; Sequence 5, Application US/09828062
; Patent No. 6710229
; GENERAL INFORMATION:
;  APPLICANT: COSTA E SILVA, OSWALDO DA
;  APPLICANT:  BOHNERT, HANS J.
;  APPLICANT:  VAN THIELEN, NOCHA
;  APPLICANT:  CHEN, ROUYING
;  APPLICANT:  SARRIA-MILLAN, RODRIGO
;  TITLE OF INVENTION: CELL CYCLE STRESS-RELATED PROTEINS AND METHODS OF USE
;  TITLE OF INVENTION:  IN PLANTS
;  FILE REFERENCE: 16313-0031
;  CURRENT APPLICATION NUMBER: US/09/828,062
;  CURRENT FILING DATE:  2001-08-20
;  PRIOR APPLICATION NUMBER: 60/196,001
;  PRIOR FILING DATE: 2000-04-07
;  NUMBER OF SEQ ID NOS: 34
;  SOFTWARE: PatentIn Ver. 2.1
; SEQ ID NO 5
;   LENGTH: 4348
;   TYPE: DNA
;   ORGANISM: Physcomitrella patens
US-09-828-062-5

  Query Match             100.0%;  Score 4348;  DB 3;  Length 4348;
  Best Local Similarity   100.0%;  Pred. No. 0;
  Matches 4348;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy          1 ATGGCGCGCCGCACTCACGTGAGGAATTGCACCTCCTTGTTCTGCGACGGTTCCATTCTT 60
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db          1 ATGGCGCGCCGCACTCACGTGAGGAATTGCACCTCCTTGTTCTGCGACGGTTCCATTCTT 60

Qy         61 TTTGGTTTTTAGTTTGCAAATCTTGATCGTGGAGTTGAGAAAAAGGGCGGTTCGTTGTCT 120
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db         61 TTTGGTTTTTAGTTTGCAAATCTTGATCGTGGAGTTGAGAAAAAGGGCGGTTCGTTGTCT 120

Qy        121 TGAGGTGTTCTTGTTGATTGTTCGTCATGGAAAATAATGATGCACTTGACATTGGAGCCG 180
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        121 TGAGGTGTTCTTGTTGATTGTTCGTCATGGAAAATAATGATGCACTTGACATTGGAGCCG 180

Qy        181 TGTCGTCCCCATATCCTTCGCAATCTGAAGGAGTGTCTACGCCATTGCCGCAAGTAACAT 240
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        181 TGTCGTCCCCATATCCTTCGCAATCTGAAGGAGTGTCTACGCCATTGCCGCAAGTAACAT 240

Qy        241 CACCGAGCTTCGACAATGCAGCCTCACCCGTGGCCGGGCGGAGGGCCGTACGGCAGACCC 300
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
```

```
Db        241 CACCGAGCTTCGACAATGCAGCCTCACCCGTGGCCGGGCGGAGGGCCGTACGGCAGACCC 300

Qy        301 CTACATCTGCAGTTCGAAGGAGAGGGAGAGAAACGGATTCCGCTCGTCGTAGGAGGAGTC 360
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        301 CTACATCTGCAGTTCGAAGGAGAGGGAGAGAAACGGATTCCGCTCGTCGTAGGAGGAGTC 360

Qy        361 GATCTCGCAGTTTAGGCAATTCTGTTTATAGTTCCCCTTACGATGCGGGGACTCCTGGAA 420
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        361 GATCTCGCAGTTTAGGCAATTCTGTTTATAGTTCCCCTTACGATGCGGGGACTCCTGGAA 420

Qy        421 CTCCTGGAACTCCAGTGGCTACTCCGGTTTACGCTACCCCAGTCGGTACACCTATGGGTA 480
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        421 CTCCTGGAACTCCAGTGGCTACTCCGGTTTACGCTACCCCAGTCGGTACACCTATGGGTA 480

Qy        481 CCCCATCGTTCCATCGTGGCACGCCACAGTACAAACAGCGCAGTGAGCTTGGTTCCCAGG 540
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        481 CCCCATCGTTCCATCGTGGCACGCCACAGTACAAACAGCGCAGTGAGCTTGGTTCCCAGG 540

Qy        541 GGAAGCCTCTACATCGGAGACGTCGATCTCAATCCAGAGAACCCGGGCATCGATCTCCTT 600
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        541 GGAAGCCTCTACATCGGAGACGTCGATCTCAATCCAGAGAACCCGGGCATCGATCTCCTT 600

Qy        601 CAAGGGAACCTAGTGCTGATGGGCGTCCCTCTGAATCTGCTGAGCCAGATGACACTTTGG 660
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        601 CAAGGGAACCTAGTGCTGATGGGCGTCCCTCTGAATCTGCTGAGCCAGATGACACTTTGG 660

Qy        661 GTGGAGAATATGCTTATGTTTGGGGACGAATGTTAACATTCCAGATGTGCTTAGGGCGA 720
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        661 GTGGAGAATATGCTTATGTTTGGGGACGAATGTTAACATTCCAGATGTGCTTAGGGCGA 720

Qy        721 TTCGTCGATTTCTCCACAATTATCGTTCGAGTGCTCATGATCTTAATTCCAAGTACATCC 780
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        721 TTCGTCGATTTCTCCACAATTATCGTTCGAGTGCTCATGATCTTAATTCCAAGTACATCC 780

Qy        781 AGATCATAGAGGAGACTGTGGAGCGTGAGGAGGATACTCTAAATATCGACATGTCAGACA 840
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        781 AGATCATAGAGGAGACTGTGGAGCGTGAGGAGGATACTCTAAATATCGACATGTCAGACA 840

Qy        841 TTTATGACCATGATCCTGATCTATACGCAAAAATTGTTCGATACCCACTCGACATCATCC 900
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        841 TTTATGACCATGATCCTGATCTATACGCAAAAATTGTTCGATACCCACTCGACATCATCC 900

Qy        901 CCCTGTTGGACACTGAGTGTCAGGAAGTTGCTACCTCTTTACTACCAACGTTTGAGAAGC 960
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        901 CCCTGTTGGACACTGAGTGTCAGGAAGTTGCTACCTCTTTACTACCAACGTTTGAGAAGC 960

Qy        961 ATATTGAGGCCAGACCTTTCAATCTCAAAGCATCGGTGCACATGCGTGAACTCAACCCTT 1020
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        961 ATATTGAGGCCAGACCTTTCAATCTCAAAGCATCGGTGCACATGCGTGAACTCAACCCTT 1020

Qy       1021 CAGATATAGACAAATTGGTTTCTGTTAAAGGAATGGTTATCCGGTGCAGTTCTATCATAC 1080
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1021 CAGATATAGACAAATTGGTTTCTGTTAAAGGAATGGTTATCCGGTGCAGTTCTATCATAC 1080

Qy       1081 CTGAAATTAAGGGGGCCTTCTTCAAATGTTTAGTGTGTGGTCACTCGCCTCCGCTAGTTA 1140
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1081 CTGAAATTAAGGGGGCCTTCTTCAAATGTTTAGTGTGTGGTCACTCGCCTCCGCTAGTTA 1140

Qy       1141 CAGTTGTTAAAGGGCGGGTTGAGGAGCCAACAAGGTGTGAAAAGCCAGAATGTGCAGCAC 1200
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1141 CAGTTGTTAAAGGGCGGGTTGAGGAGCCAACAAGGTGTGAAAAGCCAGAATGTGCAGCAC 1200

Qy       1201 GGAATGCTATGTCTCTTATTCACAATCGATGCACTTTTGCAAATAAGCAGATAGTGCGTC 1260
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1201 GGAATGCTATGTCTCTTATTCACAATCGATGCACTTTTGCAAATAAGCAGATAGTGCGTC 1260

Qy       1261 TTCAAGAAACTCCAGATGCCATTCCTGAAGGAGAGACTCCACACACAGTCAGCATGTGTT 1320
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1261 TTCAAGAAACTCCAGATGCCATTCCTGAAGGAGAGACTCCACACACAGTCAGCATGTGTT 1320

Qy       1321 TATACAACACTATGGTTGATGCTGTGAAGCCTGGAGATCGTATTGAGGTAACAGGAGTTT 1380
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1321 TATACAACACTATGGTTGATGCTGTGAAGCCTGGAGATCGTATTGAGGTAACAGGAGTTT 1380

Qy       1381 TCAAGGCCATGGCAGTTCGAGTTGGTCATGGCGCGCCGCACTCACGTGAGGAATTGCACC 1440
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1381 TCAAGGCCATGGCAGTTCGAGTTGGTCATGGCGCGCCGCACTCACGTGAGGAATTGCACC 1440

Qy       1441 TCCTTGTTCTGCGACGGTTCCATTCTTTTTGGTTTTAGTTTGCAAATCTTGATCGTGGA 1500
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1441 TCCTTGTTCTGCGACGGTTCCATTCTTTTTGGTTTTAGTTTGCAAATCTTGATCGTGGA 1500

Qy       1501 GTTGAGAAAAAGGGCGGTTCGTTGTCTTGAGGTGTTCTTGTTGATTGTTCGTCATGGAAA 1560
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1501 GTTGAGAAAAAGGGCGGTTCGTTGTCTTGAGGTGTTCTTGTTGATTGTTCGTCATGGAAA 1560

Qy       1561 ATAATGATGCACTTGACATTGGAGCCGTGTCGTCCCCATATCCTTCGCAATCTGAAGGAG 1620
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1561 ATAATGATGCACTTGACATTGGAGCCGTGTCGTCCCCATATCCTTCGCAATCTGAAGGAG 1620

Qy       1621 TGTCTACGCCATTGCCGCAAGTAACATCACCGAGCTTCGACAATGCAGCCTCACCCGTGG 1680
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1621 TGTCTACGCCATTGCCGCAAGTAACATCACCGAGCTTCGACAATGCAGCCTCACCCGTGG 1680

Qy       1681 CCGGGCGGAGGGCCGTACGGCAGACCCCTACATCTGCAGTTCGAAGGAGAGGGAGAGAAA 1740
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1681 CCGGGCGGAGGGCCGTACGGCAGACCCCTACATCTGCAGTTCGAAGGAGAGGGAGAGAAA 1740

Qy       1741 CGGATTCCGCTCGTCGTAGGAGGAGTCGATCTCGCAGTTTAGGCAATTCTGTTTATAGTT 1800
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1741 CGGATTCCGCTCGTCGTAGGAGGAGTCGATCTCGCAGTTTAGGCAATTCTGTTTATAGTT 1800

Qy       1801 CCCCTTACGATGCGGGGACTCCTGGAACTCCTGGAACTCCAGTGGCTACTCCGGTTTACG 1860
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1801 CCCCTTACGATGCGGGGACTCCTGGAACTCCTGGAACTCCAGTGGCTACTCCGGTTTACG 1860

Qy       1861 CTACCCCAGTCGGTACACCTATGGGTACCCCATCGTTCCATCGTGGCACGCCACAGTACA 1920
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1861 CTACCCCAGTCGGTACACCTATGGGTACCCCATCGTTCCATCGTGGCACGCCACAGTACA 1920

Qy       1921 AACAGCGCAGTGAGCTTGGTTCCCAGGGGAAGCCTCTACATCGGAGACGTCGATCTCAAT 1980
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1921 AACAGCGCAGTGAGCTTGGTTCCCAGGGGAAGCCTCTACATCGGAGACGTCGATCTCAAT 1980

Qy       1981 CCAGAGAACCCGGGCATCGATCTCCTTCAAGGGAACCTAGTGCTGATGGGCGTCCCTCTG 2040
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1981 CCAGAGAACCCGGGCATCGATCTCCTTCAAGGGAACCTAGTGCTGATGGGCGTCCCTCTG 2040

Qy       2041 AATCTGCTGAGCCAGATGACACTTTGGGTGGAGAATATGCTTATGTTTGGGGACGAATG 2100
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       2041 AATCTGCTGAGCCAGATGACACTTTGGGTGGAGAATATGCTTATGTTTGGGGACGAATG 2100

Qy       2101 TTAACATTCCAGATGTGCTTAGGGCGATTCGTCGATTTCTCCACAATTATCGTTCGAGTG 2160
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       2101 TTAACATTCCAGATGTGCTTAGGGCGATTCGTCGATTTCTCCACAATTATCGTTCGAGTG 2160

Qy       2161 CTCATGATCTTAATTCCAAGTACATCCAGATCATAGAGGAGACTGTGGAGCGTGAGGAGG 2220
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       2161 CTCATGATCTTAATTCCAAGTACATCCAGATCATAGAGGAGACTGTGGAGCGTGAGGAGG 2220

Qy       2221 ATACTCTAAATATCGACATGTCAGACATTTATGACCATGATCCTGATCTATACGCAAAAA 2280
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       2221 ATACTCTAAATATCGACATGTCAGACATTTATGACCATGATCCTGATCTATACGCAAAAA 2280

Qy       2281 TTGTTCGATACCCACTCGACATCATCCCCCTGTTGGACACTGAGTGTCAGGAAGTTGCTA 2340
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       2281 TTGTTCGATACCCACTCGACATCATCCCCCTGTTGGACACTGAGTGTCAGGAAGTTGCTA 2340

Qy       2341 CCTCTTTACTACCAACGTTTGAGAAGCATATTGAGGCCAGACCTTTCAATCTCAAAGCAT 2400
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       2341 CCTCTTTACTACCAACGTTTGAGAAGCATATTGAGGCCAGACCTTTCAATCTCAAAGCAT 2400

Qy       2401 CGGTGCACATGCGTGAACTCAACCCTTCAGATATAGACAAATTGGTTTCTGTTAAAGGAA 2460
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       2401 CGGTGCACATGCGTGAACTCAACCCTTCAGATATAGACAAATTGGTTTCTGTTAAAGGAA 2460
```

```
Qy       2461 TGGTTATCCGGTGCAGTTCTATCATACCTGAAATTAAGGGGGCCTTCTTCAAATGTTTAG 2520
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       2461 TGGTTATCCGGTGCAGTTCTATCATACCTGAAATTAAGGGGGCCTTCTTCAAATGTTTAG 2520

Qy       2521 TGTGTGGTCACTCGCCTCCGCTAGTTACAGTTGTTAAAGGGCGGGTTGAGGAGCCAACAA 2580
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       2521 TGTGTGGTCACTCGCCTCCGCTAGTTACAGTTGTTAAAGGGCGGGTTGAGGAGCCAACAA 2580

Qy       2581 GGTGTGAAAAGCCAGAATGTGCAGCACGGAATGCTATGTCTCTTATTCACAATCGATGCA 2640
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       2581 GGTGTGAAAAGCCAGAATGTGCAGCACGGAATGCTATGTCTCTTATTCACAATCGATGCA 2640

Qy       2641 CTTTTGCAAATAAGCAGATAGTGCGTCTTCAAGAAACTCCAGATGCCATTCCTGAAGGAG 2700
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       2641 CTTTTGCAAATAAGCAGATAGTGCGTCTTCAAGAAACTCCAGATGCCATTCCTGAAGGAG 2700

Qy       2701 AGACTCCACACACAGTCAGCATGTGTTTATACAACACTATGGTTGATGCTGTGAAGCCTG 2760
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       2701 AGACTCCACACACAGTCAGCATGTGTTTATACAACACTATGGTTGATGCTGTGAAGCCTG 2760

Qy       2761 GAGATCGTATTGAGGTAACAGGAGTTTTCAAGGCCATGGCAGTTCGAGTTGGTCCGAATC 2820
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       2761 GAGATCGTATTGAGGTAACAGGAGTTTTCAAGGCCATGGCAGTTCGAGTTGGTCCGAATC 2820

Qy       2821 AACGAACATTACGAGCATTGTATAAGACCTACATCGATTGCGTGCACGTCAAGAAGTCTG 2880
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       2821 AACGAACATTACGAGCATTGTATAAGACCTACATCGATTGCGTGCACGTCAAGAAGTCTG 2880

Qy       2881 ACAGGGGTCGACTGCAAACTGAAGATCCTATGGAGATGGATAAGGAGAATGATATGTATG 2940
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       2881 ACAGGGGTCGACTGCAAACTGAAGATCCTATGGAGATGGATAAGGAGAATGATATGTATG 2940

Qy       2941 CTGGGTATCATGAAAGTGATACTTCAGAAGCTGCTAATGAAGCAAAGATTCAAAAACTTA 3000
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       2941 CTGGGTATCATGAAAGTGATACTTCAGAAGCTGCTAATGAAGCAAAGATTCAAAAACTTA 3000

Qy       3001 AAGAGCTGTCCAAGCTCCCGGGCATTTATGATAGACTTTCAAGGTCGCTGGCTCCAAGCA 3060
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       3001 AAGAGCTGTCCAAGCTCCCGGGCATTTATGATAGACTTTCAAGGTCGCTGGCTCCAAGCA 3060

Qy       3061 TTTGGGAGCTTGAAGATATTAAAAAGGGTCTTCTTTGCCAGCTCTTTGGTGGAAGGCTA 3120
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       3061 TTTGGGAGCTTGAAGATATTAAAAAGGGTCTTCTTTGCCAGCTCTTTGGTGGAAGGCTA 3120

Qy       3121 AGAAAATTCCATCTGGAGCATCTTTCCGAGGTGACATCAATGTTTTACTTGTTGGGGACC 3180
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       3121 AGAAAATTCCATCTGGAGCATCTTTCCGAGGTGACATCAATGTTTTACTTGTTGGGGACC 3180

Qy       3181 CTGGTACCAGTAAATCTCAGCTGCTTCAGTATGTGCACAAGATAGCTCCTCGTGGAATCT 3240
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       3181 CTGGTACCAGTAAATCTCAGCTGCTTCAGTATGTGCACAAGATAGCTCCTCGTGGAATCT 3240

Qy       3241 ACACTAGTGGGCGAGGAAGTTCGGCGGTTGGGCTGACAGCGTATGTAACGAAGGATCCAG 3300
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       3241 ACACTAGTGGGCGAGGAAGTTCGGCGGTTGGGCTGACAGCGTATGTAACGAAGGATCCAG 3300

Qy       3301 AAACTCGAGAGACGGTATTGGAGAGCGGAGCTTTGGTTCTTAGTGATCGTGGGATATGCT 3360
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       3301 AAACTCGAGAGACGGTATTGGAGAGCGGAGCTTTGGTTCTTAGTGATCGTGGGATATGCT 3360

Qy       3361 GTATCGATGAGTTCGACAAAATGTCTGATAATGCCCGAAGCATGCTTCATGAGGTAATGG 3420
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       3361 GTATCGATGAGTTCGACAAAATGTCTGATAATGCCCGAAGCATGCTTCATGAGGTAATGG 3420

Qy       3421 AGCAACAAACGGTATCTGTAGCCAAAGGGGGTATCATTGCCTCGCTGAACGCTCGGACGT 3480
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       3421 AGCAACAAACGGTATCTGTAGCCAAAGGGGGTATCATTGCCTCGCTGAACGCTCGGACGT 3480

Qy       3481 CTGTCCTTGCATGTGCAAATCCTAGTGGGTCCCGATACAATGCGCGCCTTTCTGTGATTG 3540
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       3481 CTGTCCTTGCATGTGCAAATCCTAGTGGGTCCCGATACAATGCGCGCCTTTCTGTGATTG 3540

Qy       3541 ATAACATCCAGCTTCCTCCAACTCTACTTTCTAGATTTGATTTAATTTACTTAATGCTCG 3600
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       3541 ATAACATCCAGCTTCCTCCAACTCTACTTTCTAGATTTGATTTAATTTACTTAATGCTCG 3600

Qy       3601 ACAAACCAGACGAGCAAAACGATCGTCGTCTCGCCAGGCATCTCGTGGCTTTACACTATG 3660
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       3601 ACAAACCAGACGAGCAAAACGATCGTCGTCTCGCCAGGCATCTCGTGGCTTTACACTATG 3660

Qy       3661 AAAACTATGAAGTTTCAAAGCAGGACGCCTTAGATCTACAAACACTTACCGCGTATATCA 3720
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       3661 AAAACTATGAAGTTTCAAAGCAGGACGCCTTAGATCTACAAACACTTACCGCGTATATCA 3720

Qy       3721 CCTATGCTCGTCAGCATGTACATCCTACATTAAGTGATGAAGCTGCTGAAGATTTGATTA 3780
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       3721 CCTATGCTCGTCAGCATGTACATCCTACATTAAGTGATGAAGCTGCTGAAGATTTGATTA 3780

Qy       3781 ATGGCTATGTTGAGATGCGCCAAAAGGGCAACTTTCCTGGAAGCAGTAAAAAGGTGATAA 3840
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       3781 ATGGCTATGTTGAGATGCGCCAAAAGGGCAACTTTCCTGGAAGCAGTAAAAAGGTGATAA 3840

Qy       3841 CAGCCACACCTCGGCAACTCGAAAGTATGATTCGTATCAGTGAAGCCCTAGCTCGAATGA 3900
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       3841 CAGCCACACCTCGGCAACTCGAAAGTATGATTCGTATCAGTGAAGCCCTAGCTCGAATGA 3900

Qy       3901 GATTTTCTGAAGTGGTAGAGAAAGTTGATGCAGCAGAAGCTGTGCGCCTTTTAGACGTCG 3960
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       3901 GATTTTCTGAAGTGGTAGAGAAAGTTGATGCAGCAGAAGCTGTGCGCCTTTTAGACGTCG 3960

Qy       3961 CTTTGCAGCAATCTGCTACTGATCATGCAACAGGTACGATAGACATGGATCTTATCACGA 4020
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       3961 CTTTGCAGCAATCTGCTACTGATCATGCAACAGGTACGATAGACATGGATCTTATCACGA 4020

Qy       4021 CTGGAGTGTCGGCCAGCGAGCGTATTCGTCGGGCCAACTTGCTAGCTGCTCTGCGAGAGC 4080
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       4021 CTGGAGTGTCGGCCAGCGAGCGTATTCGTCGGGCCAACTTGCTAGCTGCTCTGCGAGAGC 4080

Qy       4081 TTATAGCAGATAAAATTTCACCTGGCAGCTCCTCTGGCTTGAAGACCAGTCAGCTTCTTG 4140
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       4081 TTATAGCAGATAAAATTTCACCTGGCAGCTCCTCTGGCTTGAAGACCAGTCAGCTTCTTG 4140

Qy       4141 AGGATATCCGGAGCCAAAGCAGTGTGGACGTTAGTTTGCAGGATATTAAAAATGCTCTGG 4200
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       4141 AGGATATCCGGAGCCAAAGCAGTGTGGACGTTAGTTTGCAGGATATTAAAAATGCTCTGG 4200

Qy       4201 GTAGCCTCCAAGGAGAAGGCTTTCTTACTGTCCATGGTGACATAGTCAAGAGAGTTTGAG 4260
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       4201 GTAGCCTCCAAGGAGAAGGCTTTCTTACTGTCCATGGTGACATAGTCAAGAGAGTTTGAG 4260

Qy       4261 ACAGTTTCTAACTGTTCGAATCCATGAGCTATAACTCTGAACGAAAGGAAAACCTCCAG 4320
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       4261 ACAGTTTCTAACTGTTCGAATCCATGAGCTATAACTCTGAACGAAAGGAAAACCTCCAG 4320

Qy       4321 TTTCCCATGCGCAATTCCCAGAGCTCGC 4348
              ||||||||||||||||||||||||||||
Db       4321 TTTCCCATGCGCAATTCCCAGAGCTCGC 4348
```

```
RESULT 2
US-09-828-062-2
; Sequence 2, Application US/09828062
; Patent No. 6710229
; GENERAL INFORMATION:
;  APPLICANT: COSTA E SILVA, OSWALDO DA
;  APPLICANT:  BOHNERT, HANS J.
;  APPLICANT:  VAN THIELEN, NOCHA
;  APPLICANT:  CHEN, ROUYING
;  APPLICANT:  SARRIA-MILLAN, RODRIGO
;  TITLE OF INVENTION: CELL CYCLE STRESS-RELATED PROTEINS AND METHODS OF USE
;  TITLE OF INVENTION:  IN PLANTS
;  FILE REFERENCE: 16313-0031
;  CURRENT APPLICATION NUMBER: US/09/828,062
;  CURRENT FILING DATE:  2001-08-20
;  PRIOR APPLICATION NUMBER: 60/196,001
;  PRIOR FILING DATE: 2000-04-07
```

```
;  NUMBER OF SEQ ID NOS: 34
;  SOFTWARE: PatentIn Ver. 2.1
; SEQ ID NO 2
;   LENGTH: 723
;   TYPE: DNA
;   ORGANISM: Physcomitrella patens
US-09-828-062-2

  Query Match             15.5%;  Score 673;  DB 3;  Length 723;
  Best Local Similarity   97.1%;  Pred. No. 2.2e-209;
  Matches  696;  Conservative    0;  Mismatches   20;  Indels    1;  Gaps    1;

Qy       2753 GAAGCCTGGAGATCGTATTGAGGTAACAGGAGTTTTCAAGGCCATGGCAGTTCGAGTTGG 2812
Db          3 GCACCAGGGAGATCGTATTGAGGTAACAGGAGTTTTCAAGGCCATGGCAGTTCGAGTTGG 62

Qy       2813 TCCGAATCAACGAACATTACGAGCATTGTATAAGACCTACATCGATTGCGTGCACGTCAA 2872
Db         63 TCCGAATCAACGAACATTACGAGCATTGTATAAGACCTACATCGATTGCGTGCACGTCAA 122

Qy       2873 GAAGTCTGACAGGGGTCGACTGCAAACTGAAGATCCTATGGAGATGGATAAGGAGAATGA 2932
Db        123 GAAGTCTGACAGGGGTCGACTGCAAACTGAAGATCCTATGGAGATGGATAAGGAGAATGA 182

Qy       2933 TATGTATGCTGGGTATCATGAAAGTGATACTTCAGAAGCTGCTAATGAAGCAAAGATTCA 2992
Db        183 TATGTATGCTGGGTATCATGAAAGTGATACTTCAGAAGCTGCTAATGAAGCAAAGATTCA 242

Qy       2993 AAAACTTAAAGAGCTGTCCAAGCTCCCGGGCATTTATGATAGACTTTCAAGGTCGCTGGC 3052
Db        243 AAAACTTAAAGAGCTGTCCAAGCTCCCGGACATTTATGATAGACTTTCAAGGTCGCTGGC 302

Qy       3053 TCCAAGCATTTGGGAGCTTGAAGATATTAAAAAGGGTCTTCTTTGCCAGCTCTTTGGTGG 3112
Db        303 TCCAAGCATTTGGGAGCTTGAAGATATTAAAAAGGGTCTTCTTTGCCAGCTCTTTGGTGG 362

Qy       3113 GAAGGCTAAGAAAATTCCATCTGGAGCATCTTTCCGAGGTGACATCAATGTTTTACTTGT 3172
Db        363 GAAGGCTAAGAAAATTCCATCTGGAGCATCTTTCCGAGGTGACATCAATGTTTTACTTGT 422

Qy       3173 TGGGGACCCTGGTACCAGTAAATCTCAGCTGCTTCAGTATGTGCACAAGATAGCTCCTCG 3232
Db        423 TGGGGACCCTGGTACCAGTAAATCTCAGCTGCTTCAGTATGTGCACAAGATAGCTCCTCG 482

Qy       3233 TGGAATCTACACTAGTGGGCGAGGAAGTTCGGCGGTTGGGCTGACAGCGTATGT-AACGA 3291
Db        483 TGGAATCTACACTAGTGGGCGAGGAAGTTCGGCGGTTGGGCTGACAGCGTATGTAAACGA 542

Qy       3292 AGGATCCAGAAACTCGAGAGACGGTATTGGAGAGCGGAGCTTTGGTTCTTAGTGATCGTG 3351
Db        543 AGGATCCAGAAACTCGAGAGACGGTATTGGAGAGCGGAGCTTTGGTTCTTAGTGATCGTG 602

Qy       3352 GGATATGCTGTATCGATGAGTTCGACAAAATGTCTGATAATGCCCGAAGCATGCTTCATG 3411
Db        603 GGATATGCTGTATCGATGAGTTCGACAAAATGTCTGATAATGCCCGAAGCATGCTTCATG 662

Qy       3412 AGGTAATGGAGCAACAAACGGTATCTGTAGCCAAAGGGGGTATCATTGCCTCGCTGA 3468
Db        663 AGGTAATGGAGCAACAAACGGTATCTGGACCCAAGCGGTTCATGCTCGTGAAGCCGA 719


RESULT 3
US-09-604-605-1
; Sequence 1, Application US/09604605
; Patent No. 6421613
; GENERAL INFORMATION:
;  APPLICANT:  Ramgopal Nadimpalli
;  APPLICANT:  Carl R. Simmons
;  TITLE OF INVENTION:  Maize Prolifera Gene and Uses Thereof
;  FILE REFERENCE:  1138
;  CURRENT APPLICATION NUMBER: US/09/604,605
;  CURRENT FILING DATE:  2000-06-27
;  PRIOR APPLICATION NUMBER:  60/143,222
;  PRIOR FILING DATE:  1999-07-09
;  NUMBER OF SEQ ID NOS:  2
;  SOFTWARE:  FastSEQ for Windows Version 3.0
; SEQ ID NO 1
;   LENGTH: 2533
;   TYPE: DNA
;   ORGANISM: Zea mays
;   FEATURE:
;   NAME/KEY: CDS
;   LOCATION: (110)...(2269)
US-09-604-605-1

  Query Match              4.8%;  Score 209.6;  DB 3;  Length 2533;
  Best Local Similarity   52.0%;  Pred. No. 4.2e-57;
  Matches  498;  Conservative    0;  Mismatches  454;  Indels    6;  Gaps    1;

Qy       2992 AAAAACTTAAAGAGCTGTCCAAGCTCCCGGGCATTTATGATAGACTTTCAAGGTCGCTGG 3051
Db       1059 AACAAATTGACCGATTGGCTGAGGACGGTGATATCTACAGTAAGCTGGCAAGGTCCTTGG 1118

Qy       3052 CTCCAAGCATTTGGGAGCTTGAAGATATTAAAAAGGGTCTTCTTTGCCAGCTCTTTGGTG 3111
Db       1119 CACCTGAAATATTTGGCCATGAAGATGTCAAAAAGGCACTGCTGTTACTACTTGTTGGTG 1178

Qy       3112 GGAAGGCTAAGAAAATTCCATCTGGAGCATCTTTCCGAGGTGACATCAATGTTTTACTTG 3171
Db       1179 CACCCCATCGGAAGCTCGCAGATGGCATGAAGATCAGAGGAGACCTGCATATATGCATGA 1238

Qy       3172 TTGGGGACCCTGGTACCAGTAAATCTCAGCTGCTTCAGTATGTGCACAAGATAGCTCCTC 3231
Db       1239 TGGGAGATCCTGGTGTTGCAAAGAGTCAACTTCTGAAGCATATTATCAATGTTGCACCAA 1298

Qy       3232 GTGGAATCTACACTAGTGGGCGAGGAAGTTCGGCGGTTGGGCTGACAGCGTATGTAACGA 3291
Db       1299 GAGGAGTGTACACCACTGGACGTGGGAGCAGTGGTGTTGGACTTACCGCTGCAGTCCAGA 1358

Qy       3292 AGGATCCAGAAACTCGAGAGACGGTATTGGAGAGCGGAGCTTTGGTTCTTAGTGATCGTG 3351
Db       1359 AAGATCCAGTAACAAATGAGTTTGTCCTCGAAGGTGGAGCACTGGTACTGGCAGATATGG 1418

Qy       3352 GGATATGCTGTATCGATGAGTTCGACAAAATGTCTGATAATGCCCGAAGCATGCTTCATG 3411
Db       1419 GCATTTGTGCTATAGATGAGTTTGACAAGATGGAAGAGTCAGACAGGACAGCGATTCATG 1478

Qy       3412 AGGTAATGGAGCAACAAACGGTATCTGTAGCCAAAGGGGGTATCATTGCCTCGCTGAACG 3471
Db       1479 AGGTAATGGAGCAGCAAACAGTTAGCATTGCCAAGGCTGGCATCACCACCTCTCTTAATG 1538

Qy       3472 CTCGGACGTCTGTCCTTGCATGTGCAAATCCTAGTGGGTCCCGATACAATGCGCGCCTTT 3531
Db       1539 CGAGAACTGCAATTCTGGCTGCTGCAAATCCAGCATGGGGAAGGTATGATATGAGGAGAA 1598

Qy       3532 CTGTGATTGATAACATCCAGCTTCCTCCAACTCTACTTTCTAGATTTGATTTAATTTACT 3591
Db       1599 CCCCAGCAGAAAATATAAATCTACCTCCAGCTCTTCTGTCTCGTTTCGACCTCCTTTGGT 1658

Qy       3592 TAATGCTCGACAAACCAGACGAGCAAAACGATCGTCGTCTCGCCAGGCATCTCGTGGCTT 3651
Db       1659 TAATCCTGGATCGTGCAGACATGGAAACTGATCTTGAAATGGCAAGACACGTTGTTCATG 1718

Qy       3652 TACACTATGAAAACTATGAAGTTTCAAAGCAGGACGCCTTAGATCTACAAACACTTACCG 3711
Db       1719 TGCATCAAAATCTTGAATCACCAGCGCTGGGGTTCACACCACTTGAGCCATCTGTACTCA 1778

Qy       3712 CGTATATCACCTATGCTCGTCAGCATGTACATCCTACATTAAGTGATGAAGCTGCTGAAG 3771
Db       1779 GAGCATACATATCTGCTGCTAGAAGAGTCATTCCTTCTGTTCCTAGAGAGCTCGAGGAAT 1838

Qy       3772 ATTTGATTAATGGCTATGTTGAGATGCGCCAAAAGGGCAACTTTCCTGGAAGCAGTAAAA 3831
Db       1839 ACATTGCAACTGCATATTCCAGCATCCGCCAAGAGGAGGCGAAGTCAAATGCACCAACCT 1898

Qy       3832 AGGTGATAACAGCCACACCTCGGCAACTCGAAAGTATGATTCGTATCAGTGAAGCCCTAG 3891
```

```
Db       1899 CCTACACAACTATCAGAAC------ACTTTTGAGCATACTCCGTATCTCTATTGCCTTGG 1952

Qy       3892 CTCGAATGAGATTTTCTGAAGTGGTAGAGAAAGTTGATGCAGCAGAAGCTGTGCGCCT 3949

Db       1953 CAAGACTGAGGTTCTCAGAAACTGTGGCTCAGAGCGATGTCGACGAAGCACTGCGACT 2010
```

RESULT 4
US-10-101-080-1
; Sequence 1, Application US/10101080
; Patent No. 6879916
; GENERAL INFORMATION:
; APPLICANT: Ramgopal Nadimpalli
; APPLICANT: Carl R. Simmons
; TITLE OF INVENTION: Maize Prolifera Gene and Uses Thereof
; FILE REFERENCE: 1138
; CURRENT APPLICATION NUMBER: US/10/101,080
; CURRENT FILING DATE: 2002-05-09
; PRIOR APPLICATION NUMBER: US/09/604,605
; PRIOR FILING DATE: 2000-06-27
; NUMBER OF SEQ ID NOS: 2
; SOFTWARE: FastSEQ for Windows Version 3.0
; SEQ ID NO 1
;   LENGTH: 2533
;   TYPE: DNA
;   ORGANISM: Zea mays
;   FEATURE:
;   NAME/KEY: CDS
;   LOCATION: (110)...(2269)
US-10-101-080-1

```
  Query Match             4.8%;  Score 209.6;  DB 3;  Length 2533;
  Best Local Similarity   52.0%;  Pred. No. 4.2e-57;
  Matches  498;  Conservative    0;  Mismatches  454;  Indels    6;  Gaps    1;

Qy       2992 AAAAACTTAAAGAGCTGTCCAAGCTCCCGGGCATTTATGATAGACTTTCAAGGTCGCTGG 3051

Db       1059 AACAAATTGACCGATTGGCTGAGGACGGTGATATCTACAGTAAGCTGGCAAGGTCCTTGG 1118

Qy       3052 CTCCAAGCATTTGGGAGCTTGAAGATATTAAAAAGGGTCTTCTTTGCCAGCTCTTTGGTG 3111

Db       1119 CACCTGAAATATTTGGCCATGAAGATGTCAAAAAGGCACTGCTGTTACTACTTGTTGGTG 1178

Qy       3112 GGAAGGCTAAGAAAATTCCATCTGGAGCATCTTTCCGAGGTGACATCAATGTTTTACTTG 3171

Db       1179 CACCCCATCGGAAGCTCGCAGATGGCATGAAGATCAGAGGAGACCTGCATATATGCATGA 1238

Qy       3172 TTGGGGACCCTGGTACCAGTAAATCTCAGCTGCTTCAGTATGTGCACAAGATAGCTCCTC 3231

Db       1239 TGGGAGATCCTGGTGTTGCAAAGAGTCAACTTCTGAAGCATATTATCAATGTTGCACCAA 1298

Qy       3232 GTGGAATCTACACTAGTGGGCGAGGAAGTTCGGCGGTTGGGCTGACAGCGTATGTAACGA 3291

Db       1299 GAGGAGTGTACACCACTGGACGTGGGAGCAGTGGTGTTGGACTTACCGCTGCAGTCCAGA 1358

Qy       3292 AGGATCCAGAAACTCGAGAGACGGTATTGGAGAGCGGAGCTTTGGTTCTTAGTGATCGTG 3351

Db       1359 AAGATCCAGTAACAAATGAGTTTGTCCTCGAAGGTGGAGCACTGGTACTGGCAGATATGG 1418

Qy       3352 GGATATGCTGTATCGATGAGTTCGACAAAATGTCTGATAATGCCCGAAGCATGCTTCATG 3411

Db       1419 GCATTTGTGCTATAGATGAGTTTGACAAGATGGAAGAGTCAGACAGGACAGCGATTCATG 1478

Qy       3412 AGGTAATGGAGCAACAAACGGTATCTGTAGCCAAAGGGGGTATCATTGCCTCGCTGAACG 3471

Db       1479 AGGTAATGGAGCAGCAAACAGTTAGCATTGCCAAGGCTGGCATCACCACCTCTCTTAATG 1538

Qy       3472 CTCGGACGTCTGTCCTTGCATGTGCAAATCCTAGTGGGTCCCGATACAATGCGCGCCTTT 3531

Db       1539 CGAGAACTGCAATTCTGGCTGCTGCAAATCCAGCATGGGGAAGGTATGATATGAGGAGAA 1598

Qy       3532 CTGTGATTGATAACATCCAGCTTCCTCCAACTCTACTTTCTAGATTTGATTTAATTTACT 3591

Db       1599 CCCCAGCAGAAAATATAAATCTACCTCCAGCTCTTCTGTCTCGTTTCGACCTCCTTTGGT 1658

Qy       3592 TAATGCTCGACAAACCAGACGAGCAAAACGATCGTCGTCTCGCCAGGCATCTCGTGGCTT 3651

Db       1659 TAATCCTGGATCGTGCAGACATGGAAACTGATCTTGAAATGGCAAGACACGTTGTTCATG 1718

Qy       3652 TACACTATGAAAACTATGAAGTTTCAAAGCAGGACGCCTTAGATCTACAAACACTTACCG 3711

Db       1719 TGCATCAAAATCTTGAATCACCAGCGCTGGGGTTCACACCACTTGAGCCATCTGTACTCA 1778

Qy       3712 CGTATATCACCTATGCTCGTCAGCATGTACATCCTACATTAAGTGATGAAGCTGCTGAAG 3771

Db       1779 GAGCATACATATCTGCTGCTAGAAGAGTCATTCCTTCTGTTCCTAGAGAGCTCGAGGAAT 1838

Qy       3772 ATTTGATTAATGGCTATGTTGAGATGCGCCAAAAGGGCAACTTTCCTGGAAGCAGTAAAA 3831

Db       1839 ACATTGCAACTGCATATTCCAGCATCCGCCAAGAGGAGGCGAAGTCAAATGCACCAACCT 1898

Qy       3832 AGGTGATAACAGCCACACCTCGGCAACTCGAAAGTATGATTCGTATCAGTGAAGCCCTAG 3891

Db       1899 CCTACACAACTATCAGAAC------ACTTTTGAGCATACTCCGTATCTCTATTGCCTTGG 1952

Qy       3892 CTCGAATGAGATTTTCTGAAGTGGTAGAGAAAGTTGATGCAGCAGAAGCTGTGCGCCT 3949

Db       1953 CAAGACTGAGGTTCTCAGAAACTGTGGCTCAGAGCGATGTCGACGAAGCACTGCGACT 2010
```

RESULT 5
US-09-248-796A-4534
; Sequence 4534, Application US/09248796A
; Patent No. 6747137
; GENERAL INFORMATION:
; APPLICANT: Keith Weinstock et al
; TITLE OF INVENTION: NUCLEIC ACID AND AMINO ACID SEQUENCES RELATING TO CANDIDA ALBICA
; TITLE OF INVENTION: FOR DIAGNOSTICS AND THERAPEUTICS
; FILE REFERENCE: 107196.132
; CURRENT APPLICATION NUMBER: US/09/248,796A
; CURRENT FILING DATE: 1999-02-12
; PRIOR APPLICATION NUMBER: US 60/074,725
; PRIOR FILING DATE: 1998-02-13
; PRIOR APPLICATION NUMBER: US 60/096,409
; PRIOR FILING DATE: 1998-08-13
; NUMBER OF SEQ ID NOS: 28208
; SEQ ID NO 4534
;   LENGTH: 1635
;   TYPE: DNA
;   ORGANISM: Candida albicans
US-09-248-796A-4534

```
  Query Match             4.6%;  Score 198.6;  DB 3;  Length 1635;
  Best Local Similarity   55.8%;  Pred. No. 1.2e-53;
  Matches  378;  Conservative    0;  Mismatches  299;  Indels    0;  Gaps    0;

Qy       2979 GAAGCAAAGATTCAAAAACTTAAAGAGCTGTCCAAGCTCCCGGGCATTTATGATAGACTT 3038

Db        445 GAAGAAGAAATACGTGAATTTAGAAAATTATCACACGAAAAAGGTATTATTGATAAAATC 504

Qy       3039 TCAAGGTCGCTGGCTCCAAGCATTTGGGAGCTTGAAGATATTAAAAAGGGTCTTCTTTGC 3098

Db        505 ATTGCCTCAATGGCTCCATCCATATATGGTCATAAAGATATTAAAACTGCCATTGCTTGT 564

Qy       3099 CAGCTCTTTGGTGGGAAGGCTAAGAAAATTCCATCTGGAGCATCTTTCCGAGGTGACATC 3158

Db        565 TCATTATTTGGTGGTGTGCCTAAAGATGTTAATGGAAACTTTCCATTAGAGGTGATATT 624

Qy       3159 AATGTTTTACTTGTTGGGGACCCTGGTACCAGTAAATCTCAGCTGCTTCAGTATGTGCAC 3218

Db        625 AATGTACTATTATTAGGGGATCCAGGTACTGCTAAATCGCAAATATTGAAATATGCTGAG 684

Qy       3219 AAGATAGCTCCTCGTGGAATCTACACTAGTGGGCGAGGAAGTTCGGCGGTTGGGCTGACA 3278

Db        685 AAAACTGCTAGTCGTGCTGTTTTTGCAACTGGTCAAGGTGCCTCGGCTGTTGGTCTTACA 744
```

```
Qy       3279 GCGTATGTAACGAAGGATCCAGAAACTCGAGAGACGGTATTGGAGAGCGGAGCTTTGGTT 3338
Db        745 GCATCGGTTAGAAAAGACCCAATCACTCGTGAATGGACATTACAAGGTGGGGCTTTGGTG 804

Qy       3339 CTTAGTGATCGTGGGATATGCTGTATCGATGAGTTCGACAAAATGTCTGATAATGCCCGA 3398
Db        805 CTTGCAGATAAAGGGACTTGTCTTATTGATGAATTTGATAAGATGAATGATCAAGATAGA 864

Qy       3399 AGCATGCTTCATGAGGTAATGGAGCAACAAACGGTATCTGTAGCCAAAGGGGGTATCATT 3458
Db        865 ACTTCTATTCATGAAGCTATGGAACAACAAAGTATATCTATTTCTAAAGCTGGTATCGTC 924

Qy       3459 GCCTCGCTGAACGCTCGGACGTCTGTCCTTGCATGTGCAAATCCTAGTGGGTCCCGATAC 3518
Db        925 ACTACCTTACATGCCAGATGTGCTGTTATCGCTGCTGCCAATCCTAATGGAGGTCGATAC 984

Qy       3519 AATGCGCGCCTTTCTGTGATTGATAACATCCAGCTTCCTCCAACTCTACTTTCTAGATTT 3578
Db        985 AACTCTACATTACCATTACCAGAAAATGTCGATTTAACTTACCCAATTTTATCCAGATTT 1044

Qy       3579 GATTTAATTTACTTAATGCTCGACAAACCAGACGAGCAAAACGATCGTCGTCTCGCCAGG 3638
Db       1045 GATATTATGTGTATCGTGCGAGATTTGGTCAATCCTGAATCTGATGAACGATTAGCTAGT 1104

Qy       3639 CATCTCGTGGCTTTACA 3655
Db       1105 TTTGTCATTGATTCCCA 1121
```

RESULT 6
US-09-248-796A-5904
; Sequence 5904, Application US/09248796A
; Patent No. 6747137
; GENERAL INFORMATION:
; APPLICANT: Keith Weinstock et al
; TITLE OF INVENTION: NUCLEIC ACID AND AMINO ACID SEQUENCES RELATING TO CANDIDA ALBICAN
; TITLE OF INVENTION: FOR DIAGNOSTICS AND THERAPEUTICS
; FILE REFERENCE: 107196.132
; CURRENT APPLICATION NUMBER: US/09/248,796A
; CURRENT FILING DATE: 1999-02-12
; PRIOR APPLICATION NUMBER: US 60/074,725
; PRIOR FILING DATE: 1998-02-13
; PRIOR APPLICATION NUMBER: US 60/096,409
; PRIOR FILING DATE: 1998-08-13
; NUMBER OF SEQ ID NOS: 28208
; SEQ ID NO 5904
;   LENGTH: 1488
;   TYPE: DNA
;   ORGANISM: Candida albicans
;   FEATURE:
;   NAME/KEY: unsure
;   LOCATION: (2)
;   OTHER INFORMATION: Identity of nucleotide sequences at the above locations are unknc
US-09-248-796A-5904

```
  Query Match             4.5%;  Score 197.6;  DB 3;  Length 1488;
  Best Local Similarity   51.3%;  Pred. No. 2.5e-53;
  Matches  514;  Conservative    0;  Mismatches  479;  Indels    9;  Gaps    2;

Qy       2939 TGCTGGGTATCATGAAAGTGATACTTCAGAAGCTGCTAATGAAGCAAAGATTCAAAAACT 2998
Db        144 TGATTTGAATGACCAAGAAGTTTTTCTTACATCTTTGAGTGATGCTGAAGTATTACAATT 203

Qy       2999 TAAAGAGCTGTCCAAGCTCCCGGGCATTTATGATAGACTTTCAAGGTCGCTGGCTCCAAG 3058
Db        204 AAAGGAAATGGTCAAGGATGAACATATTTATGACAAGCTTGTGAACTCAATTGCACCAGC 263

Qy       3059 CATTTGGGAGCTTGAAGATATTAAAAAGGGTCTTCTTTGCCAGCTCTTTGGTGGGAAGGC 3118
Db        264 TGTGTTTGGCCACGAAGTGATCAAAAAGGGAATTCTTTTGCAATTATTAGGCGGTGTTCA 323

Qy       3119 TAAGAAAATTCCATCTGGAGCATCTTTCCGAGGTGACATCAATGTTTTACTTGTTGGGGA 3178
Db        324 CAAACAAACAGTAGATGGAATCAACTTAAGAGGGGATATAAACATTTGTATTGTTGGTGA 383

Qy       3179 CCCTGGTACCAGTAAATCTCAGCTGCTTCAGTATGTGCACAAGATAGCTCCTCGTGGAAT 3238
Db        384 CCCATCCACTTCGAAATCACAGTTTTGAAGTATGTCTGTGGGTTTTCGCCTCGTGCTGT 443

Qy       3239 CTACACTAGTGGGCGAGGAAGTTCGGCGGTTGGCTGACAGCGTATGTAACGAAGGATCC 3298
Db        444 CTACACTTCTGGTAAGGCTTCTTCTGCTGCTGGGTTGACAGCGGCTGTAGTGAAGGACGA 503

Qy       3299 AGAAACTCGAGAGACGGTATTGGAGAGCGGAGCTTTGGTTCTTAGTGATCGTGGGATATG 3358
Db        504 AGAAAGCGGAGAATACACTATTGAAGCAGGTGCTTTGATGTTGGCAGATAACGGTATTTG 563

Qy       3359 CTGTATCGATGAGTTCGACAAAATGTCTGATAATGCCCGAAGCATGCTTCATGAGGTAAT 3418
Db        564 TGCAATTGATGAATTTGACAAGATGGATATTGCTGACCAAGTTGCTATTCACGAAGCCAT 623

Qy       3419 GGAGCAACAAACGGTATCTGTAGCCAAAGGGGGTATCATTGCCTCGCTGAACGCTCGGAC 3478
Db        624 GGAACAACAAACAATCTCCATTGCAAAAGCTGGTATTCATGCTACCTTAAATGCCCGTAC 683

Qy       3479 GTCTGTCCTTGCATGTGCAAATCCTAGTGGGTCCCGATACAATGCGCGCCTTTCTGTGAT 3538
Db        684 ATCTATCTTAGCAGCTGCTAACCCTATTGGGGGTAGATATAATAGAAAATTGGGTCTTCG 743

Qy       3539 TGATAACATCCAGCTTCCTCCAACTCTACTTTCTAGATTTGATTTAATTTACTTAATGCT 3598
Db        744 TTCTAATTTGAATATGACAGCACCAATTATGTCGAGATTTGATTTGTTCTTTGTTGTTCT 803

Qy       3599 CGACAAACCAGACGAGCAAAACGATCGTCGTCTCGCCAGGCATCTCGTGGCTTTACACTA 3658
Db        804 TGATGATTGTAATGAAAGAATTGATACTCAATTAGCATCACACATTGTTGATTTGCACAT 863

Qy       3659 TGAAAACTATGAAGTTTCAAAGCAGGACGCCTTAGATCTACAAACACTTACCGCGTATAT 3718
Db        864 GCTTAGAGACGATGCAATTGATCCACCTTACTCAGCCGAACAATTA------GCAAGATA 917

Qy       3719 CACCTATGCTCGTCAGCATGTACATCCTACATTAAGTGATGAAGCTGCTGAAGATTTGAT 3778
Db        918 CATTAAATACGCAAAGACTTTCAAGCCTAGAATGACTAAAGAAGCTAGAGACTTTTTGGT 977

Qy       3779 TAATGGCTATGTTGAGATGCGCCAAAAGGGCAACTTTCCTGGAAGCAGTAAAAAGGTGAT 3838
Db        978 CACTAGATACAAGGAATTGAGAGAAGATG---ATGCTCAAGGTTTAGGGAGGTCCTCGTA 1034

Qy       3839 AACAGCCACACCTCGGCAACTCGAAAGTATGATTCGTATCAGTGAAGCCCTAGCTCGAAT 3898
Db       1035 CAGAATTACTGTTAGACAATTGGAGTCCATGATTAGATTGTCAGAAGCTATAGCCAGAGC 1094

Qy       3899 GAGATTTTCTGAAGTGGTAGAGAAAGTTGATGCAGCAGAAGC 3940
Db       1095 CAATTGTACCGAAGAAATAACGCCAAGTTTTGTTGCTGAGGC 1136
```

RESULT 7
US-09-949-016-2468
; Sequence 2468, Application US/09949016
; Patent No. 6812339
; GENERAL INFORMATION:
; APPLICANT: VENTER, J. Craig et al.
; TITLE OF INVENTION: POLYMORPHISMS IN KNOWN GENES ASSOCIATED
; TITLE OF INVENTION: WITH HUMAN DISEASE, METHODS OF DETECTION AND USES THEREOF
; FILE REFERENCE: CL001307
; CURRENT APPLICATION NUMBER: US/09/949,016
; CURRENT FILING DATE: 2000-04-14
; PRIOR APPLICATION NUMBER: 60/241,755
; PRIOR FILING DATE: 2000-10-20
; PRIOR APPLICATION NUMBER: 60/237,768
; PRIOR FILING DATE: 2000-10-03
; PRIOR APPLICATION NUMBER: 60/231,498
; PRIOR FILING DATE: 2000-09-08

```
;  NUMBER OF SEQ ID NOS: 207012
;  SOFTWARE: FastSEQ for Windows Version 4.0
; SEQ ID NO 2468
;   LENGTH: 2917
;   TYPE: DNA
;   ORGANISM: Human
US-09-949-016-2468

  Query Match             4.3%;  Score 186.4;  DB 3;  Length 2917;
  Best Local Similarity   51.7%;  Pred. No. 1.9e-49;
  Matches  477;  Conservative    0;  Mismatches  436;  Indels    9;  Gaps    2;

Qy       3025 TTTATGATAGACTTTCAAGGTCGCTGGCTCCAAGCATTTGGGAGCTTGAAGATATTAAAA 3084
Db       1098 TATACCACAATCTTTGTACCAGCCTGTTCCCTACTATACATGGCAATGATGAAGTAAAAC 1157

Qy       3085 AGGGTCTTCTTTGCCAGCTCTTTGGTGGGAAGGCTAAGAAAATTCCATCTGGAGCATCTT 3144
Db       1158 GGGGTGTCCTGCTGATGCTCTTTGGTGGCGTTCCAAAGACAACAGGAGAAGGGACCTCTC 1217

Qy       3145 TCCGAGGTGACATCAATGTTTTACTTGTTGGGGACCCTGGTACCAGTAAATCTCAGCTGC 3204
Db       1218 TTCGAGGGGACATAAATGTTTGCATTGTTGGTGACCCAAGTACAGCTAAGAGCCAATTTC 1277

Qy       3205 TTCAGTATGTGCACAAGATAGCTCCTCGTGGAATCTACACTAGTGGGCGAGGAAGTTCGG 3264
Db       1278 TCAAGCACGTGGAGGAGTTCAGCCCCAGAGCTGTCTACACCAGTGGTAAAGCGTCCAGTG 1337

Qy       3265 CGGTTGGGCTGACAGCGTATGTAACGAAGGATCCAGAAACTCGAGAGACGGTATTGGAGA 3324
Db       1338 CTGCTGGCTTAACAGCAGCTGTTGTGAGAGATGAAGAATCTCATGAGTTTGTCATTGAGG 1397

Qy       3325 GCGGAGCTTTGGTTCTTAGTGATCGTGGGATATGCTGTATCGATGAGTTCGACAAAATGT 3384
Db       1398 CTGGAGCTTTGATGTTGGCTGATAATGGTGTGTGTTGTATTGATGAATTTGATAAGATGG 1457

Qy       3385 CTGATAATGCCCGAAGCATGCTTCATGAGGTAATGGAGCAACAAACGGTATCTGTAGCCA 3444
Db       1458 ACGTGCGGGATCAAGTTGCTATTCATGAAGCTATGGAACAGCAGACCATATCCATCACTA 1517

Qy       3445 AAGGGGGTATCATTGCCTCGCTGAACGCTCGGACGTCTGTCCTTGCATGTGCAAATCCTA 3504
Db       1518 AAGCAGGAGTGAAGGCTACTCTGAACGCCCGGACGTCCATTTTGGCAGCAGCAAACCCAA 1577

Qy       3505 GTGGGTCCCGATACAATGCGCGCCTTTCTGTGATTGATAACATCCAGCTTCCTCCAACTC 3564
Db       1578 TCAGTGGACACTATGACAGATCAAAATCATTGAAACAGAATATAAATTTGTCAGCTCCCA 1637

Qy       3565 TACTTTCTAGATTTGATTTAATTTACTTAATGCTCGACAAACCAGACGAGCAAAACGATC 3624
Db       1638 TCATGTCCCGATTCGATCTCTTCTTTATCCTTGTGGATGAATGTAATGAGGTTACAGATT 1697

Qy       3625 GTCGTCTCGCCAGGCATCTCGTGGCTTTACACTATGAAAACTATGAAGTTTCAAAGCAGG 3684
Db       1698 ATGCCATTGCCAGGCGCATAGTAGATTTGC---ATTCAAGAATTGAGGAATCAATTGATC 1754

Qy       3685 ACGCCTTAGATCTACAAACACTTACCGCGTATATCACCTATGCTCGTCAGCATGTACATC 3744
Db       1755 GTGTCTATTCCCTCGATGATATCAGAAGATATCTTCTCTTTGCAAGACAGTTTAAAC--- 1811

Qy       3745 CTACATTAAGTGATGAAGCTGCTGAAGATTTGATTAATGGCTATGTTGAGATGCGCCAAA 3804
Db       1812 ---CCAAGATTTCCAAAGAGTCAGAGGACTTCATTGTGGAGCAATATAAACATCTCCGCC 1868

Qy       3805 AGGGCAACTTTCCTGGAAGCAGTAAAAAGGTGATAACAGCCACACCTCGGCAACTCGAAA 3864
Db       1869 AGAGAGATGGTTCTGGAGTGACCAAGTCTTCATGGAGGATTACAGTGCGACAGCTTGAGA 1928

Qy       3865 GTATGATTCGTATCAGTGAAGCCCTAGCTCGAATGAGATTTTCTGAAGTGGTAGAGAAAG 3924
Db       1929 GCATGATTCGTCTCTCTGAAGCTATGGCTCGGATGCACTGCTGTGATGAGGTCCAACCTA 1988

Qy       3925 TTGATGCAGCAGAAGCTGTGCG 3946
Db       1989 AACATGTGAAGGAAGCTTTCCG 2010

RESULT 8
US-09-949-016-5021
; Sequence 5021, Application US/09949016
; Patent No. 6812339
; GENERAL INFORMATION:
;  APPLICANT: VENTER, J. Craig et al.
;  TITLE OF INVENTION: POLYMORPHISMS IN KNOWN GENES ASSOCIATED
;  TITLE OF INVENTION:  WITH HUMAN DISEASE, METHODS OF DETECTION AND USES THEREOF
;  FILE REFERENCE: CL001307
;  CURRENT APPLICATION NUMBER: US/09/949,016
;  CURRENT FILING DATE:  2000-04-14
;  PRIOR APPLICATION NUMBER: 60/241,755
;  PRIOR FILING DATE: 2000-10-20
;  PRIOR APPLICATION NUMBER: 60/237,768
;  PRIOR FILING DATE: 2000-10-03
;  PRIOR APPLICATION NUMBER: 60/231,498
;  PRIOR FILING DATE: 2000-09-08
;  NUMBER OF SEQ ID NOS: 207012
;  SOFTWARE: FastSEQ for Windows Version 4.0
; SEQ ID NO 5021
;   LENGTH: 2533
;   TYPE: DNA
;   ORGANISM: Human
US-09-949-016-5021

  Query Match             4.2%;  Score 183.8;  DB 3;  Length 2533;
  Best Local Similarity   56.2%;  Pred. No. 1.2e-48;
  Matches  366;  Conservative    0;  Mismatches  282;  Indels    3;  Gaps    1;

Qy       3006 CTGTCCAAGCTCCCGGGCATTTATGATAGACTTTCAAGGTCGCTGGCTCCAAGCATTTGG 3065
Db       1055 CTGGCTGCCCTCCCAAATGTCTATGAGGTCATCTCCAAGAGCATCGCCCCCTCCATCTTT 1114

Qy       3066 GAGCTTGAAGATATTAAAAAGGGTCTTCTTTGCCAGCTCTTTGGTGGGAAGGCTAAGAAA 3125
Db       1115 GGGGGCACAGACATGAAGAAGGCCATTGCCTGCCTGCTCTTTGGGGGCTCCCGAAAGAGG 1174

Qy       3126 ATTCCATCTGGAGCATCTTTCCGAGGTGACATCAATGTTTTACTTGTTGGGGACCCTGGT 3185
Db       1175 CTCCCTGATGGACTTACTCGCCGAGGAGACATCAACCTGCTGATGCTAGGGGACCCTGGG 1234

Qy       3186 ACCAGTAAATCTCAGCTGCTTCAGTATGTGCACAAGATAGCTCCTCGTGGAATCTACACT 3245
Db       1235 ACAGCCAAGTCCCAGCTTCTGAAGTTTGTGGAGAAGTGTTCTCCCATTGGGGTATACACG 1294

Qy       3246 AGTGGGCGAGGAAGTTCGGCGGTTGGGCTGACAGCGTATGTAACGAAGGATCCAGAAACT 3305
Db       1295 TCTGGGAAAGGCAGCAGCGCAGCTGGACTGACAGCCTCGGTGATGAGGGACCCTTCGTCC 1354

Qy       3306 CGAGAGACGGTATTGGAGAGCGGAGCTTTGGTTCTTAGTGATCGTGGGATATGCTGTATC 3365
Db       1355 CGGAATTTCATCATGGAGGGCGGAGCCATGGTCCTGGCCGATGGTGGGGTCGTCTGTATT 1414

Qy       3366 GATGAGTTCGACAAAATGTCTGATAATGCCCGAAGCATGCTTCATGAGGTAATGGAGCAA 3425
Db       1415 GACGAGTTTGACAAGATGCGAGAAGATGACCGTGTGGCAATCCACGAAGCCATGGAGCAG 1474

Qy       3426 CAAACGGTATCTGTAGCCAAAGGGGGTATCATTGCCTCGCTGAACGCTCGGACGTCTGTC 3485
Db       1475 CAGACCATCTCTATCGCCAAGGCTGGGATCACCACCACCCTGAACTCCCGCTGCTCCGTC 1534

Qy       3486 CTTGCATGTGCAAATCCTAGTGGGTCCCGATACAATGCGCGCCTTTCTGTGATTGATAAC 3545
Db       1535 CTGGCTGCTGCCAACTCAGTGTTCGGCCGCTGGGATGAG---ACGAAGGGGGAGGACAAC 1591

Qy       3546 ATCCAGCTTCCTCCAACTCTACTTTCTAGATTTGATTTAATTTACTTAATGCTCGACAAA 3605
Db       1592 ATTGACTTCATGCCCACCATCTTGTCGCGCTTCGACATGATCTTCATCGTCAAGGATGAG 1651
```

```
Qy      3606 CCAGACGAGCAAAACGATCGTCGTCTCGCCAGGCATCTCGTGGCTTTACAC 3656
Db      1652 CACAATGAGGAGAGGGATGTGATGCTGGCCAAGCATGTCATCACTCTGCAC 1702
```

RESULT 9
US-09-949-016-2621
; Sequence 2621, Application US/09949016
; Patent No. 6812339
; GENERAL INFORMATION:
; APPLICANT: VENTER, J. Craig et al.
; TITLE OF INVENTION: POLYMORPHISMS IN KNOWN GENES ASSOCIATED
; TITLE OF INVENTION: WITH HUMAN DISEASE, METHODS OF DETECTION AND USES THEREOF
; FILE REFERENCE: CL001307
; CURRENT APPLICATION NUMBER: US/09/949,016
; CURRENT FILING DATE: 2000-04-14
; PRIOR APPLICATION NUMBER: 60/241,755
; PRIOR FILING DATE: 2000-10-20
; PRIOR APPLICATION NUMBER: 60/237,768
; PRIOR FILING DATE: 2000-10-03
; PRIOR APPLICATION NUMBER: 60/231,498
; PRIOR FILING DATE: 2000-09-08
; NUMBER OF SEQ ID NOS: 207012
; SOFTWARE: FastSEQ for Windows Version 4.0
; SEQ ID NO 2621
; LENGTH: 3000
; TYPE: DNA
; ORGANISM: Human
US-09-949-016-2621

```
  Query Match             4.1%;  Score 180;  DB 3;  Length 3000;
  Best Local Similarity   51.2%;  Pred. No. 2.4e-47;
  Matches  479;  Conservative    0;  Mismatches  445;  Indels   12;  Gaps    2;

Qy      3020 GGGCATTTATGATAGACTTTCAAGGTCGCTGGCTCCAAGCATTTGGGAGCTTGAAGATAT 3079
Db      1625 GGATTTCTACGAAAAGCTGGCAGCTTCAATCGCCCCAGAAATATACGGGCATGAAGATGT 1684

Qy      3080 TAAAAAGGGTCTTCTTTGCCAGCTCTTTGGTGGGAAGGCTAAGAAAATTCCATCTGGAGC 3139
Db      1685 GAAGAAGGCACTGCTGCTCCTGCTAGTCGGGGGTGTGG---ACCAGTCTCCTCGAGGCAT 1741

Qy      3140 ATCTTTCCGAGGTGACATCAATGTTTTACTTGTTGGGGACCCTGGTACCAGTAAATCTCA 3199
Db      1742 GAAAATCCGGGGCAACATCAACATCTGTCTGATGGGGATCCTGGTGTGGCCAAGTCTCA 1801

Qy      3200 GCTGCTTCAGTATGTGCACAAGATAGCTCCTCGTGGAATCTACACTAGTGGGCGAGGAAG 3259
Db      1802 GCTCCTGTCATACATTGATCGACTGGCGCCTCGCAGCCAGTACACAACAGGCCGGGGCTC 1861

Qy      3260 TTCGGCGGTTGGGCTGACAGCGTATGTAACGAAGGATCCAGAAACTCGAGAGACGGTATT 3319
Db      1862 CTCAGGAGTGGGGCTTACGGCAGCTGTGCTGAGAGACTCCGTGAGTGGAGAACTGACCTT 1921

Qy      3320 GGAGAGCGGAGCTTTGGTTCTTAGTGATCGTGGGATATGCTGTATCGATGAGTTCGACAA 3379
Db      1922 AGAGGGTGGGGCCCTGGTGCTGGCTGACCAGGGTGTGTGCTGCATTGATGAGTTCGACAA 1981

Qy      3380 AATGTCTGATAATGCCCGAAGCATGCTTCATGAGGTAATGGAGCAACAAACGGTATCTGT 3439
Db      1982 GATGGCTGAGGCCGACCGCACAGCCATCCACGAGGTCATGGAGCAGCAGACCATCTCCAT 2041

Qy      3440 AGCCAAAGGGGGTATCATTGCCTCGCTGAACGCTCGGACGTCTGTCCTTGCATGTGCAAA 3499
Db      2042 TGCCAAGGCCGGCATTCTCACCACACTCAATGCCCGCTGCTCCATCCTGGCTGCCGCCAA 2101

Qy      3500 TCCTAGTGGGTCCCGATACAATGCGCGCCTTTCTGTGATTGATAACATCCAGCTTCCTCC 3559
Db      2102 CCCTGCCTACGGGCGCTACAACCCTCGCCGCAGCCTGGAGCAGAACATACAGCTACCTGC 2161

Qy      3560 AACTCTACTTTCTAGATTTGATTTAATTTACTTAATGCTCGACAAACCAGACGAGCAAAA 3619
Db      2162 TGCACTGCTCTCCCGGTTTGACCTCCTCTGGCTGATTCAGGACCGGCCCGACCGAGACAA 2221

Qy      3620 CGATCGTCGTCTCGCCAGGCATCTCGTGGCTTTACACTATGAAAACTATGAAGTTTCAAA 3679
Db      2222 TGACCTACGGTTGGCCCAGCACATCACCTATGTGCACCAGCACAGCCGGCAGCCCCCCTC 2281

Qy      3680 GCAGGACGCCTTAGATCTACAAACACTTACCGCGTATATCACCTATGCTCGTCAGCATGT 3739
Db      2282 CCAGTTTGAACCTCTGGACATGAAGCTCATGAGGCGTTACATAGCCATGTGCCGCGAGAA 2341

Qy      3740 ACATCCTACATTAAGTGATGAAGCTGCTGAAGATTTGATTAATGGCTATGTTGAGATGCG 3799
Db      2342 GCAGCCCATGGTGCCAGAGTCTCTGGCTGACTACATCACAGCAGCATACGTGGAGATGAG 2401

Qy      3800 CCAAAAGGGCAACTTTCCTGGAAGCAGTAAAAAGGTGATAACAGCCACACCTCGGCAACT 3859
Db      2402 GCGAGAGG---------CTTGGGCTAGTAAGGATGCCACCTATACTTCTGCCCGGACCCT 2452

Qy      3860 CGAAAGTATGATTCGTATCAGTGAAGCCCTAGCTCGAATGAGATTTTCTGAAGTGGTAGA 3919
Db      2453 GCTGGCTATCCTGCGCCTTTCCACTGCTCTGGCACGTCTGAGAATGGTGGATGTGGTGGA 2512

Qy      3920 GAAAGTTGATGCAGCAGAAGCTGTGCGCCTTTTAGA 3955
Db      2513 GAAAGAAGATGTGAATGAAGCCATCAGGCTAATGGA 2548
```

RESULT 10
US-09-248-796A-4570
; Sequence 4570, Application US/09248796A
; Patent No. 6747137
; GENERAL INFORMATION:
; APPLICANT: Keith Weinstock et al
; TITLE OF INVENTION: NUCLEIC ACID AND AMINO ACID SEQUENCES RELATING TO CANDIDA ALBICAN
; TITLE OF INVENTION: FOR DIAGNOSTICS AND THERAPEUTICS
; FILE REFERENCE: 107196.132
; CURRENT APPLICATION NUMBER: US/09/248,796A
; CURRENT FILING DATE: 1999-02-12
; PRIOR APPLICATION NUMBER: US 60/074,725
; PRIOR FILING DATE: 1998-02-13
; PRIOR APPLICATION NUMBER: US 60/096,409
; PRIOR FILING DATE: 1998-08-13
; NUMBER OF SEQ ID NOS: 28208
; SEQ ID NO 4570
; LENGTH: 996
; TYPE: DNA
; ORGANISM: Candida albicans
US-09-248-796A-4570

```
  Query Match             4.0%;  Score 173.2;  DB 3;  Length 996;
  Best Local Similarity   56.5%;  Pred. No. 1.9e-45;
  Matches  322;  Conservative    0;  Mismatches  248;  Indels    0;  Gaps    0;

Qy      2992 AAAAACTTAAAGAGCTGTCCAAGCTCCCGGGCATTTATGATAGACTTTCAAGGTCGCTGG 3051
Db       407 ACAAAATCGACGAGCTTTTACTTCAAGGTGATGTTTACAATAAGTTGGCCAAATCCATTG 466

Qy      3052 CTCCAAGCATTTGGGAGCTTGAAGATATTAAAAAGGGTCTTCTTTGCCAGCTCTTTGGTG 3111
Db       467 CTCCGGAAATTTATGGTCACTTGGATGTTAAGAAAATTTTATTATTGTTATTATGTGGTG 526

Qy      3112 GGAAGGCTAAGAAAATTCCATCTGGAGCATCTTTCCGAGGTGACATCAATGTTTTACTTG 3171
Db       527 GTGTCACCAAAGAAATCGGTGATGGGTTGAAAATTAGAGGTGATATTAATGTTTGTCTTA 586

Qy      3172 TTGGGGACCCTGGTACCAGTAAATCTCAGCTGCTTCAGTATGTGCACAAGATAGCTCCTC 3231
Db       587 TGGGTGATCCGGGGGTTGCTAAATCACAATTATTGAAAGCCATTGGTAAGATTGCTCCAA 646

Qy      3232 GTGGAATCTACACTAGTGGGCGAGGAAGTTCGGCGGTTGGGCTGACAGCGTATGTAACGA 3291
Db       647 GATCGGTCTATACTACTGGTAGAGGTTCTTCTGGTGTAGGGTTGACGGCAGCTGTTATGA 706

Qy      3292 AGGATCCAGAAACTCGAGAGACGGTATTGGAGAGCGGAGCTTTGGTTCTTAGTGATCGTG 3351
```

```
Db        707 GAGATCCAATTACTGATGAAATGGTGTTGGAAGGTGGGGCTTTGGTGTTGGCCGATAACG 766

Qy       3352 GGATATGCTGTATCGATGAGTTCGACAAAATGTCTGATAATGCCCGAAGCATGCTTCATG 3411
Db        767 GTATTTGTTGTATTGATGAATTTGATAAAATGGACGAAAGCGACAGAACGGCGATTCATG 826

Qy       3412 AGGTAATGGAGCAACAAACGGTATCTGTAGCCAAAGGGGGTATCATTGCCTCGCTGAACG 3471
Db        827 AAGTTATGGAACAACAAACCATCTCTATTGCCAAAGCTGGTATCACTACTACTTTGAATG 886

Qy       3472 CTCGGACGTCTGTCCTTGCATGTGCAAATCCTAGTGGGTCCCGATACAATGCGCGCCTTT 3531
Db        887 CCAGAACTTCGATCTTGGCAGCAGCCAATCCATTATATGGTAGATATAATCCAAGATTGT 946

Qy       3532 CTGTGATTGATAACATCCAGCTTCCTCCAA 3561
Db        947 CTCCACATGAAAATATCAATTTACCAGCCA 976
```

RESULT 11
US-09-964-899-48
; Sequence 48, Application US/09964899
; Patent No. 6900367
; GENERAL INFORMATION:
; APPLICANT: Cohen, Dalia et al.
; TITLE OF INVENTION: Identification of Genes Involved in
; TITLE OF INVENTION: Alzheimer's Disease Using Drosophila Melanogaster
; FILE REFERENCE: 4-31612 A
; CURRENT APPLICATION NUMBER: US/09/964,899
; CURRENT FILING DATE: 2001-09-27
; PRIOR APPLICATION NUMBER: 60/236,893
; PRIOR FILING DATE: 2000-09-29
; PRIOR APPLICATION NUMBER: 60/298,309
; PRIOR FILING DATE: 2001-06-14
; NUMBER OF SEQ ID NOS: 53
; SOFTWARE: FastSEQ for Windows Version 4.0
; SEQ ID NO 48
;   LENGTH: 3378
;   TYPE: DNA
;   ORGANISM: Homo Sapien
US-09-964-899-48

```
  Query Match             3.7%;  Score 160.8;  DB 3;  Length 3378;
  Best Local Similarity   50.9%;  Pred. No. 5.1e-41;
  Matches  381;  Conservative    0;  Mismatches  367;  Indels    0;  Gaps    0;

Qy       2940 GCTGGGTATCATGAAAGTGATACTTCAGAAGCTGCTAATGAAGCAAAGATTCAAAAACTT 2999
Db       1333 GCCAAGAAGGACAACAAGGTTGCTGTAGGGGAACTGACCGATGAAGATGTGAAGATGATC 1392

Qy       3000 AAAGAGCTGTCCAAGCTCCCGGGCATTTATGATAGACTTTCAAGGTCGCTGGCTCCAAGC 3059
Db       1393 ACTAGCCTCTCCAAGGATCAGCAGATCGGAGAGAAGATCTTTGCCAGCATTGCTCCTTCC 1452

Qy       3060 ATTTGGGAGCTTGAAGATATTAAAAAGGGTCTTCTTTGCCAGCTCTTTGGTGGGAAGGCT 3119
Db       1453 ATCTATGGTCATGAAGACATCAAGAGAGGCCTGGCTCTGGCCCTGTTCGGAGGGGAGCCC 1512

Qy       3120 AAGAAAATTCCATCTGGAGCATCTTTCCGAGGTGACATCAATGTTTTACTTGTTGGGGAC 3179
Db       1513 AAAAACCCAGGTGGCAAGCACAAGGTACGTGGTGATATCAACGTGCTCTTGTGCGGAGAC 1572

Qy       3180 CCTGGTACCAGTAAATCTCAGCTGCTTCAGTATGTGCACAAGATAGCTCCTCGTGGAATC 3239
Db       1573 CCTGGCACAGCGAAGTCGCAGTTTCTCAAGTATATTGAGAAAGTGTCCAGCCGAGCCATC 1632

Qy       3240 TACACTAGTGGGCGAGGAAGTTCGGCGGTTGGGCTGACAGCGTATGTAACGAAGGATCCA 3299
Db       1633 TTCACCACTGGCCAGGGGGCGTCGGCTGTGGGCCTCACGGCGTATGTCCAGCGGCACCCT 1692

Qy       3300 GAAACTCGAGAGACGGTATTGGAGAGCGGAGCTTTGGTTCTTAGTGATCGTGGGATATGC 3359
Db       1693 GTCAGCAGGGAGTGGACCTTGGAGGCTGGGGCCCTGGTTCTGGCTGACCGAGGAGTGTGT 1752

Qy       3360 TGTATCGATGAGTTCGACAAAATGTCTGATAATGCCCGAAGCATGCTTCATGAGGTAATG 3419
Db       1753 CTCATTGATGAATTTGACAAGATGAATGACCAGGACAGAACCAGCATCCATGAGGCCATG 1812

Qy       3420 GAGCAACAAACGGTATCTGTAGCCAAAGGGGGTATCATTGCCTCGCTGAACGCTCGGACG 3479
Db       1813 GAGCAACAGAGCATCTCCATCTCGAAGGCTGGCATCGTCACCTCCCTGCAGGCTCGCTGC 1872

Qy       3480 TCTGTCCTTGCATGTGCAAATCCTAGTGGGTCCCGATACAATGCGCGCCTTTCTGTGATT 3539
Db       1873 ACGGTCATTGCTGCCGCCAACCCCATAGGAGGGCGCTACGACCCCTCGCTGACTTTCTCT 1932

Qy       3540 GATAACATCCAGCTTCCTCCAACTCTACTTTCTAGATTTGATTTAATTTACTTAATGCTC 3599
Db       1933 GAGAACGTGGACCTCACAGAGCCCATCATCTCACGCTTTGACATCCTGTGTGTGGTGAGG 1992

Qy       3600 GACAAACCAGACGAGCAAAACGATCGTCGTCTCGCCAGGCATCTCGTGGCTTTACACTAT 3659
Db       1993 GACACCGTGGACCCAGTCCAGGACGAGATGCTGGCCCGCTTCGTGGTGGGCAGCCACGTC 2052

Qy       3660 GAAAACTATGAAGTTTCAAAGCAGGACG 3687
Db       2053 AGACACCACCCCAGCAACAAGGAGGAGG 2080
```

RESULT 12
US-09-976-594-323
; Sequence 323, Application US/09976594
; Patent No. 6673549
; GENERAL INFORMATION:
; APPLICANT: Furness, Michael
; APPLICANT: Buchbinder, Jenny
; TITLE OF INVENTION: GENES EXPRESSED IN C3A LIVER CELL CULTURES TREATED WITH STEROIDS
; FILE REFERENCE: PA-0041 US
; CURRENT APPLICATION NUMBER: US/09/976,594
; CURRENT FILING DATE: 2001-10-12
; PRIOR APPLICATION NUMBER: 60/240,409
; PRIOR FILING DATE: 2000-10-12
; NUMBER OF SEQ ID NOS: 1143
; SOFTWARE: PERL Program
; SEQ ID NO 323
;   LENGTH: 3445
;   TYPE: DNA
;   ORGANISM: Homo sapiens
;   FEATURE:
;   NAME/KEY: misc_feature
;   OTHER INFORMATION: Incyte ID No. 6673549 1723834CB1
US-09-976-594-323

```
  Query Match             3.7%;  Score 160.8;  DB 3;  Length 3445;
  Best Local Similarity   50.9%;  Pred. No. 5.1e-41;
  Matches  381;  Conservative    0;  Mismatches  367;  Indels    0;  Gaps    0;

Qy       2940 GCTGGGTATCATGAAAGTGATACTTCAGAAGCTGCTAATGAAGCAAAGATTCAAAAACTT 2999
Db       1387 GCCAAGAAGGACAACAAGGTTGCTGTAGGGGAACTGACCGATGAAGATGTGAAGATGATC 1446

Qy       3000 AAAGAGCTGTCCAAGCTCCCGGGCATTTATGATAGACTTTCAAGGTCGCTGGCTCCAAGC 3059
Db       1447 ACTAGCCTCTCCAAGGATCAGCAGATCGGAGAGAAGATCTTTGCCAGCATTGCTCCTTCC 1506

Qy       3060 ATTTGGGAGCTTGAAGATATTAAAAAGGGTCTTCTTTGCCAGCTCTTTGGTGGGAAGGCT 3119
Db       1507 ATCTATGGTCATGAAGACATCAAGAGAGGCCTGGCTCTGGCCCTGTTCGGAGGGGAGCCC 1566

Qy       3120 AAGAAAATTCCATCTGGAGCATCTTTCCGAGGTGACATCAATGTTTTACTTGTTGGGGAC 3179
Db       1567 AAAAACCCAGGTGGCAAGCACAAGGTACGTGGTGATATCAACGTGCTCTTGTGCGGAGAC 1626

Qy       3180 CCTGGTACCAGTAAATCTCAGCTGCTTCAGTATGTGCACAAGATAGCTCCTCGTGGAATC 3239
```

```
Db       1627 CCTGGCACAGCGAAGTCGCAGTTTCTCAAGTATATTGAGAAAGTGTCCAGCCGAGCCATC 1686

Qy       3240 TACACTAGTGGGCGAGGAAGTTCGGCGGTTGGGCTGACAGCGTATGTAACGAAGGATCCA 3299
Db       1687 TTCACCACTGGCCAGGGGGCGTCGGCTGTGGGCCTCACGGCGTATGTCCAGCGGCACCCT 1746

Qy       3300 GAAACTCGAGAGACGGTATTGGAGAGCGGAGCTTTGGTTCTTAGTGATCGTGGGATATGC 3359
Db       1747 GTCAGCAGGGAGTGGACCTTGGAGGCTGGGGCCCTGGTTCTGGCTGACCGAGGAGTGTGT 1806

Qy       3360 TGTATCGATGAGTTCGACAAAATGTCTGATAATGCCCGAAGCATGCTTCATGAGGTAATG 3419
Db       1807 CTCATTGATGAATTTGACAAGATGAATGACCAGGACAGAACCAGCATCCATGAGGCCATG 1866

Qy       3420 GAGCAACAAACGGTATCTGTAGCCAAAGGGGGTATCATTGCCTCGCTGAACGCTCGGACG 3479
Db       1867 GAGCAACAGAGCATCTCCATCTCGAAGGCTGGCATCGTCACCTCCCTGCAGGCTCGCTGC 1926

Qy       3480 TCTGTCCTTGCATGTGCAAATCCTAGTGGGTCCCGATACAATGCGCGCCTTTCTGTGATT 3539
Db       1927 ACGGTCATTGCTGCCGCCAACCCCATAGGAGGGCGCTACGACCCCTCGCTGACTTTCTCT 1986

Qy       3540 GATAACATCCAGCTTCCTCCAACTCTACTTTCTAGATTTGATTTAATTTACTTAATGCTC 3599
Db       1987 GAGAACGTGGACCTCACAGAGCCCATCATCTCACGCTTTGACATCCTGTGTGTGGTGAGG 2046

Qy       3600 GACAAACCAGACGAGCAAAACGATCGTCGTCTCGCCAGGCATCTCGTGGCTTTACACTAT 3659
Db       2047 GACACCGTGGACCCAGTCCAGGACGAGATGCTGGCCCGCTTCGTGGTGGGCAGCCACGTC 2106

Qy       3660 GAAAACTATGAAGTTTCAAAGCAGGACG 3687
Db       2107 AGACACCACCCCAGCAACAAGGAGGAGG 2134
```

```
RESULT 13
US-09-220-132-12
; Sequence 12, Application US/09220132
; Patent No. 6506607
; GENERAL INFORMATION:
;  APPLICANT:  Shyjan, Andrew W.
;  TITLE OF INVENTION:  METHODS AND COMPOSITONS FOR THE IDENTIFICATION AND ASSESSMENT
;  TITLE OF INVENTION:  OF PROSTATE CANCER THERAPIES AND THE DIAGNOSIS OF PROSTATE CANCE
;  FILE REFERENCE:  07334-074001
;  CURRENT APPLICATION NUMBER: US/09/220,132
;  CURRENT FILING DATE:  1998-12-23
;  PRIOR APPLICATION NUMBER:  US 60/079,303
;  PRIOR FILING DATE:  1998-03-25
;  PRIOR APPLICATION NUMBER:  US 60/068,821
;  PRIOR FILING DATE:  1997-12-24
;  NUMBER OF SEQ ID NOS:  191
;  SOFTWARE:  FastSEQ for Windows Version 4.0
; SEQ ID NO 12
;   LENGTH: 3379
;   TYPE: DNA
;   ORGANISM: Homo sapiens
US-09-220-132-12

  Query Match              3.6%;  Score 157.6;  DB 3;  Length 3379;
  Best Local Similarity   50.7%;  Pred. No. 5.7e-40;
  Matches  379;  Conservative    0;  Mismatches  369;  Indels    0;  Gaps    0;

Qy       2940 GCTGGGTATCATGAAAGTGATACTTCAGAAGCTGCTAATGAAGCAAAGATTCAAAAACTT 2999
Db       1330 GCCAAGAAGGACAACAAGGTTGCTGTAGGGGAACTGACCGATGAAGATGTGAAGATGATC 1389

Qy       3000 AAAGAGCTGTCCAAGCTCCCGGGCATTTATGATAGACTTTCAAGGTCGCTGGCTCCAAGC 3059
Db       1390 ACTAGCCTCTCCAAGGATCAGCAGATCGGAGAGAAGATCTTTGCCAGCATTGCTCCTTCC 1449

Qy       3060 ATTTGGGAGCTTGAAGATATTAAAAAGGGTCTTCTTTGCCAGCTCTTTGGTGGGAAGGCT 3119
Db       1450 ATCTATGGTCATGAAGACATCAAGAGAGGCCCTGCTCTGGCCCTGTTCGGAGGGGAGCCC 1509

Qy       3120 AAGAAAATTCCATCTGGAGCATCTTTCCGAGGTGACATCAATGTTTTACTTGTTGGGGAC 3179
Db       1510 AAAAACCCAGGTGGCAAGCACAAGGTACGTGGTGATATCAACGTGCTCTTGTGCGGAGAC 1569

Qy       3180 CCTGGTACCAGTAAATCTCAGCTGCTTCAGTATGTGCACAAGATAGCTCCTCGTGGAATC 3239
Db       1570 CCTGGCACAGCGAAGTCGCAGTTTCTCAAGTATATTGAGAAAGTGTCCAGCCGAGCCATC 1629

Qy       3240 TACACTAGTGGGCGAGGAAGTTCGGCGGTTGGGCTGACAGCGTATGTAACGAAGGATCCA 3299
Db       1630 TTCACCACTGGCCAGGGGGCGTCGGCTGTGGCCGTCACGGCGTATGTCCAGCGGCACCCT 1689

Qy       3300 GAAACTCGAGAGACGGTATTGGAGAGCGGAGCTTTGGTTCTTAGTGATCGTGGGATATGC 3359
Db       1690 GTCAGCAGGGAGTGGACCTTGGAGGCTGGGGCCCTGGTTCTGGCTGACCGAGGAGTGTGT 1749

Qy       3360 TGTATCGATGAGTTCGACAAAATGTCTGATAATGCCCGAAGCATGCTTCATGAGGTAATG 3419
Db       1750 CTCATTGATGAATTTGACAAGATGAATGACCAGGACAGAACCAGCATCCATGAGGCCATG 1809

Qy       3420 GAGCAACAAACGGTATCTGTAGCCAAAGGGGGTATCATTGCCTCGCTGAACGCTCGGACG 3479
Db       1810 GAGCAACAGAGCATCTCCATCTCGAAGGCTGGCATCGTCACCTCCCTGCAGGCTCGCTGC 1869

Qy       3480 TCTGTCCTTGCATGTGCAAATCCTAGTGGGTCCCGATACAATGCGCGCCTTTCTGTGATT 3539
Db       1870 ACGGTCATTGCTGCCGCCAACCCCATAGGAGGGCGCTACGACCCCTCGCTGACTTTCTCT 1929

Qy       3540 GATAACATCCAGCTTCCTCCAACTCTACTTTCTAGATTTGATTTAATTTACTTAATGCTC 3599
Db       1930 GAGAACGTGGACCTCACAGAGCCCATCATCTCACGCTTTGACATCCTGTGTGTGGTGAGG 1989

Qy       3600 GACAAACCAGACGAGCAAAACGATCGTCGTCTCGCCAGGCATCTCGTGGCTTTACACTAT 3659
Db       1990 GACACCGTGGACCCAGTCCAGGACGAGATGCTGGCCCGCTTCGTGGTGGGCAGCCACGTC 2049

Qy       3660 GAAAACTATGAAGTTTCAAAGCAGGACG 3687
Db       2050 AGACACCACCCCAGCAACAAGGAGGAGG 2077
```

```
RESULT 14
US-09-313-294A-4610
; Sequence 4610, Application US/09313294A
; Patent No. 6476212
; GENERAL INFORMATION:
;  APPLICANT:  Lalgudi, Raghunath V.
;  APPLICANT:  Ito, Laura Y.
;  APPLICANT:  Sherman, Bradley K.
;  TITLE OF INVENTION:  POLYNUCLEOTIDES AND POLYPEPTIDES DERIVED FROM CORN EAR
;  FILE REFERENCE:  PL-0017 US
;  CURRENT APPLICATION NUMBER: US/09/313,294A
;  CURRENT FILING DATE:  1999-05-14
;  NUMBER OF SEQ ID NOS:  7600
;  SOFTWARE:  PERL Program
; SEQ ID NO 4610
;   LENGTH: 294
;   TYPE: DNA
;   ORGANISM: Zea mays
;   FEATURE:
;   NAME/KEY: misc_feature
;   OTHER INFORMATION: Incyte ID No. 6476212 700348726H1
;   NAME/KEY: unsure
;   LOCATION: 74
;   OTHER INFORMATION: a, t, c, g, or other
US-09-313-294A-4610

  Query Match              3.2%;  Score 137;  DB 3;  Length 294;
  Best Local Similarity   66.9%;  Pred. No. 5.9e-34;
  Matches  194;  Conservative    0;  Mismatches   96;  Indels    0;  Gaps    0;

Qy       3816 CCTGGAAGCAGTAAAAAGGTGATAACAGCCACACCTCGGCAACTCGAAAGTATGATTCGT 3875
```

```
Db          5 CCTGGGAGCAGAAAGAAGGTCATAACAGCAACCGCTAGACAAATAGAGAGTTTGATCCGT 64

Qy       3876 ATCAGTGAAGCCCTAGCTCGAATGAGATTTTCTGAAGTGGTAGAGAAAGTTGATGCAGCA 3935

Db         65 CTCAGCGAANATTTAGCCCGAATGCGGTTCTCTGAAGTGGTCGAGGTGCGGGATGTTGTG 124

Qy       3936 GAAGCTGTGCGCCTTTTAGACGTCGCTTTGCAGCAATCTGCTACTGATCATGCAACAGGT 3995

Db        125 GAGGCATTCAGGCTTCTTGAAGTCGCCATGCAGCAGTCTGCGACGGATCATGCAACTGGT 184

Qy       3996 ACGATAGACATGGATCTTATCACGACTGGAGTGTCGGCCAGCGAGCGTATTCGTCGGGCC 4055

Db        185 ACGATTGATATGGATCTGATCATGACGGGGATATCCGCAAGCGAAAGGCAGAGGCGGGAG 244

Qy       4056 AACTTGCTAGCTGCTCTGCGAGAGCTTATAGCAGATAAAATTTCACCTGG 4105

Db        245 AACCTCGTTGCCGCAACCCGTAAACCTGATTGCGGAGAAAATGCAGCTTG 294


RESULT 15
US-09-313-294A-83
; Sequence 83, Application US/09313294A
; Patent No. 6476212
; GENERAL INFORMATION:
;  APPLICANT:  Lalgudi, Raghunath V.
;  APPLICANT:  Ito, Laura Y.
;  APPLICANT:  Sherman, Bradley K.
;  TITLE OF INVENTION:  POLYNUCLEOTIDES AND POLYPEPTIDES DERIVED FROM CORN EAR
;  FILE REFERENCE:  PL-0017 US
;  CURRENT APPLICATION NUMBER: US/09/313,294A
;  CURRENT FILING DATE:  1999-05-14
;  NUMBER OF SEQ ID NOS:  7600
;  SOFTWARE:  PERL Program
; SEQ ID NO 83
;   LENGTH: 311
;   TYPE: DNA
;   ORGANISM: Zea mays
;   FEATURE:
;   NAME/KEY: misc_feature
;   OTHER INFORMATION: Incyte ID No. 6476212 700548481H1
US-09-313-294A-83

  Query Match             3.1%;  Score 135.6;  DB 3;  Length 311;
  Best Local Similarity   64.8%;  Pred. No. 1.8e-33;
  Matches  201;  Conservative    0;  Mismatches  109;  Indels    0;  Gaps    0;

Qy       3643 TCGTGGCTTTACACTATGAAAACTATGAAGTTTCAAAGCAGGACGCCTTAGATCTACAAA 3702

Db          2 TTGTTTCGTTGCATTTTGAGAATCCAAATTTAGAGGAGCTCGAGGTGTTGGATCTGCAGA 61

Qy       3703 CACTTACCGCGTATATCACCTATGCTCGTCAGCATGTACATCCTACATTAAGTGATGAAG 3762

Db         62 CACTAGTTTCCTACATAAGCTATGCAAGGAAGTATATTCAGCCACAGTTAACTGATGAAG 121

Qy       3763 CTGCTGAAGATTTGATTAATGGCTATGTTGAGATGCGCCAAAAGGGCAACTTTCCTGGAA 3822

Db        122 CTGCAGAGGAGTTAACTCGTGGCTATGTTGAGATGAGGAAAAGAGGGAATAGCCCTGGGA 181

Qy       3823 GCAGTAAAAAGGTGATAACAGCCACACCTCGGCAACTCGAAAGTATGATTCGTATCAGTG 3882

Db        182 GCAGAAAGAAGGTCATAACTGCAACTGCTAGACAAATAGAGAGTTTGATTCGTCTAGGTG 241

Qy       3883 AAGCCCTAGCTCGAATGAGATTTTCTGAAGTGGTAGAGAAAGTTGATGCAGCAGAAGCTG 3942

Db        242 AAGCATTAGCCCGGATGCGGTTCTCTGAAGTGGTCGAGGTGCGAGATGTTGTGGAGGCCT 301

Qy       3943 TGCGCCTTTT 3952

Db        302 TCAGGCTTCT 311
```

Search completed: December 6, 2005, 09:16:50

Job time : 513 secs

November 2005

Published_Applications Nucleic Acid and Published_Applications Amino Acid database searches now generate two sets of results each. The Published_Applications databases have been split into two parts to reduce the amount of time required for their daily updates. This results in more machine time being available for processing searches.

Newly published applications will appear in the Published_Applications_New databases: older published applications make up the Published_Applications_Main databases.

Searches run against Nucleic Acid Published_Applications produce two sets of results, with the extensions **.rnpbm** (Published_Applications_NA_Main) and **.rnpbn** (Published_Applications_NA_New).
Searches run against Amino Acid Published_Applications produce two sets of results, with the extensions **.rapbm** (Published_Applications_AA_Main) and **.rapbn** (Published_Applications_AA_New).

OM nucleic - nucleic search, using sw model

```
Run on:          December  6, 2005, 01:50:09 ; Search time 2129 Seconds
                                          (without alignments)
                                          16888.330 Million cell updates/sec

Title:           US-10-768-511-5
Perfect score:   4348
Sequence:        1 atggcgcgccgcactcacgt..........gcgcaattcccagagctcgc 4348

Scoring table:   IDENTITY_NUC
                 Gapop 10.0 , Gapext 1.0

Searched:        9793542 seqs, 4134689005 residues

Total number of hits satisfying chosen parameters:      19587084

Minimum DB seq length: 0
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%
                 Maximum Match 100%
                 Listing first 45 summaries

Database :       Published_Applications_NA_Main:*
                 1:  /cgn2_6/ptodata/1/pubpna/US07_PUBCOMB.seq:*
                 2:  /cgn2_6/ptodata/1/pubpna/US08_PUBCOMB.seq:*
                 3:  /cgn2_6/ptodata/1/pubpna/US09A_PUBCOMB.seq:*
                 4:  /cgn2_6/ptodata/1/pubpna/US09B_PUBCOMB.seq:*
                 5:  /cgn2_6/ptodata/1/pubpna/US10A_PUBCOMB.seq:*
                 6:  /cgn2_6/ptodata/1/pubpna/US10B_PUBCOMB.seq:*
                 7:  /cgn2_6/ptodata/1/pubpna/US10C_PUBCOMB.seq:*
                 8:  /cgn2_6/ptodata/1/pubpna/US10D_PUBCOMB.seq:*
                 9:  /cgn2_6/ptodata/1/pubpna/US10E_PUBCOMB.seq:*
                 10:  /cgn2_6/ptodata/1/pubpna/US11_PUBCOMB.seq:*
```

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	% Query Match	Length	DB	ID	Description
1	4348	100.0	4348	3	US-09-828-062-5	Sequence 5, Appli
2	4348	100.0	4348	7	US-10-768-511-5	Sequence 5, Appli
3	880.2	20.2	2841	7	US-10-425-114-28735	Sequence 28735, A
4	857.2	19.7	3054	7	US-10-437-963-44639	Sequence 44639, A
5	690.8	15.9	2186	8	US-10-425-115-98898	Sequence 98898, A
6	673	15.5	723	3	US-09-828-062-2	Sequence 2, Appli
7	673	15.5	723	7	US-10-768-511-2	Sequence 2, Appli
8	634.2	14.6	1571	7	US-10-425-114-19722	Sequence 19722, A
9	598.6	13.8	2939	8	US-10-425-115-98897	Sequence 98897, A
10	492.8	11.3	2970	7	US-10-320-797-2332	Sequence 2332, Ap
11	490	11.3	1723	7	US-10-260-238-580	Sequence 580, App
12	480.2	11.0	2733	6	US-10-032-585-6480	Sequence 6480, Ap
13	470.8	10.8	2802	6	US-10-369-493-46198	Sequence 46198, A
14	451.6	10.4	2625	5	US-10-128-714-2558	Sequence 2558, Ap
15	451.6	10.4	3072	5	US-10-128-714-7558	Sequence 7558, Ap
16	430.8	9.9	3273	3	US-09-954-456-568	Sequence 568, App
17	430.8	9.9	3273	3	US-09-954-456-569	Sequence 569, App
18	430.8	9.9	3273	3	US-09-954-456-976	Sequence 976, App
19	430.8	9.9	3273	6	US-10-172-118-1920	Sequence 1920, Ap
20	430.8	9.9	3273	7	US-10-342-887-1920	Sequence 1920, Ap
21	430.8	9.9	3273	9	US-10-843-641A-3595	Sequence 3595, Ap
22	430.8	9.9	3273	9	US-10-843-641A-3596	Sequence 3596, Ap
23	430.8	9.9	3273	9	US-10-843-641A-4003	Sequence 4003, Ap
24	430.8	9.9	3273	9	US-10-756-149-1918	Sequence 1918, Ap
25	430.4	9.9	3394	6	US-10-240-965-223	Sequence 223, App
26	429.2	9.9	3395	5	US-10-084-817-287	Sequence 287, App
27	419	9.6	2842	3	US-09-925-300-113	Sequence 113, App
28	390.2	9.0	1105	7	US-10-425-114-6988	Sequence 6988, Ap
29	352.4	8.1	679	8	US-10-425-115-6272	Sequence 6272, Ap
30	352	8.1	2881	6	US-10-369-493-27488	Sequence 27488, A
31	337.2	7.8	2944	5	US-10-128-714-1558	Sequence 1558, Ap
32	337.2	7.8	3362	5	US-10-128-714-6558	Sequence 6558, Ap
33	337.2	7.8	4944	5	US-10-128-714-558	Sequence 558, App
34	337.2	7.8	5362	5	US-10-128-714-5558	Sequence 5558, Ap
35	328	7.5	2692	10	US-11-097-143-1688	Sequence 1688, Ap
36	321.4	7.4	595	7	US-10-021-323-15811	Sequence 15811, A
37	314.8	7.2	560	7	US-10-424-599-64201	Sequence 64201, A
38	228	5.2	1899	5	US-10-128-714-2100	Sequence 2100, Ap
39	228	5.2	1913	5	US-10-128-714-1100	Sequence 1100, Ap
40	228	5.2	2445	5	US-10-128-714-7100	Sequence 7100, Ap
41	228	5.2	2608	5	US-10-128-714-6100	Sequence 6100, Ap
42	228	5.2	3913	5	US-10-128-714-100	Sequence 100, App
43	228	5.2	4608	5	US-10-128-714-5100	Sequence 5100, Ap
44	226.6	5.2	1638	6	US-10-369-493-46370	Sequence 46370, A
45	222.4	5.1	5065	10	US-11-097-143-1687	Sequence 1687, Ap

ALIGNMENTS

```
RESULT 1
US-09-828-062-5
; Sequence 5, Application US/09828062
; Publication No. US20030097675A1
; GENERAL INFORMATION:
;  APPLICANT: COSTA E SILVA, OSWALDO DA
;  APPLICANT:  BOHNERT, HANS J.
;  APPLICANT:  VAN THIELEN, NOCHA
;  APPLICANT:  CHEN, ROUYING
;  APPLICANT:  SARRIA-MILLAN, RODRIGO
;  TITLE OF INVENTION: CELL CYCLE STRESS-RELATED PROTEINS AND METHODS OF USE
;  TITLE OF INVENTION:  IN PLANTS
;  FILE REFERENCE: 16313-0031
;  CURRENT APPLICATION NUMBER: US/09/828,062
;  CURRENT FILING DATE:  2001-08-20
;  PRIOR APPLICATION NUMBER: 60/196,001
;  PRIOR FILING DATE: 2000-04-07
;  NUMBER OF SEQ ID NOS: 34
;  SOFTWARE: PatentIn Ver. 2.1
; SEQ ID NO 5
;   LENGTH: 4348
;   TYPE: DNA
;   ORGANISM: Physcomitrella patens
US-09-828-062-5

  Query Match             100.0%;  Score 4348;  DB 3;  Length 4348;
  Best Local Similarity   100.0%;  Pred. No. 0;
  Matches 4348;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy          1 ATGGCGCGCCGCACTCACGTGAGGAATTGCACCTCCTTGTTCTGCGACGGTTCCATTCTT 60
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db          1 ATGGCGCGCCGCACTCACGTGAGGAATTGCACCTCCTTGTTCTGCGACGGTTCCATTCTT 60

Qy         61 TTTGGTTTTTAGTTTGCAAATCTTGATCGTGGAGTTGAGAAAAAGGGCGGTTCGTTGTCT 120
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db         61 TTTGGTTTTTAGTTTGCAAATCTTGATCGTGGAGTTGAGAAAAAGGGCGGTTCGTTGTCT 120

Qy        121 TGAGGTGTTCTTGTTGATTGTTCGTCATGGAAAATAATGATGCACTTGACATTGGAGCCG 180
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        121 TGAGGTGTTCTTGTTGATTGTTCGTCATGGAAAATAATGATGCACTTGACATTGGAGCCG 180

Qy        181 TGTCGTCCCCATATCCTTCGCAATCTGAAGGAGTGTCTACGCCATTGCCGCAAGTAACAT 240
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        181 TGTCGTCCCCATATCCTTCGCAATCTGAAGGAGTGTCTACGCCATTGCCGCAAGTAACAT 240

Qy        241 CACCGAGCTTCGACAATGCAGCCTCACCCGTGGCCGGGCGGAGGGCCGTACGGCAGACCC 300
```

```
                ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        241 CACCGAGCTTCGACAATGCAGCCTCACCCGTGGCCGGGCGGAGGGCCGTACGGCAGACCC 300

Qy        301 CTACATCTGCAGTTCGAAGGAGAGGGAGAGAAACGGATTCCGCTCGTCGTAGGAGGAGTC 360
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        301 CTACATCTGCAGTTCGAAGGAGAGGGAGAGAAACGGATTCCGCTCGTCGTAGGAGGAGTC 360

Qy        361 GATCTCGCAGTTTAGGCAATTCTGTTTATAGTTCCCCTTACGATGCGGGGACTCCTGGAA 420
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        361 GATCTCGCAGTTTAGGCAATTCTGTTTATAGTTCCCCTTACGATGCGGGGACTCCTGGAA 420

Qy        421 CTCCTGGAACTCCAGTGGCTACTCCGGTTTACGCTACCCCAGTCGGTACACCTATGGGTA 480
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        421 CTCCTGGAACTCCAGTGGCTACTCCGGTTTACGCTACCCCAGTCGGTACACCTATGGGTA 480

Qy        481 CCCCATCGTTCCATCGTGGCACGCCACAGTACAAACAGCGCAGTGAGCTTGGTTCCCAGG 540
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        481 CCCCATCGTTCCATCGTGGCACGCCACAGTACAAACAGCGCAGTGAGCTTGGTTCCCAGG 540

Qy        541 GGAAGCCTCTACATCGGAGACGTCGATCTCAATCCAGAGAACCCGGGCATCGATCTCCTT 600
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        541 GGAAGCCTCTACATCGGAGACGTCGATCTCAATCCAGAGAACCCGGGCATCGATCTCCTT 600

Qy        601 CAAGGGAACCTAGTGCTGATGGGCGTCCCTCTGAATCTGCTGAGCCAGATGACACTTTGG 660
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        601 CAAGGGAACCTAGTGCTGATGGGCGTCCCTCTGAATCTGCTGAGCCAGATGACACTTTGG 660

Qy        661 GTGGAGAATATGCTTATGTTTGGGGGACGAATGTTAACATTCCAGATGTGCTTAGGGCGA 720
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        661 GTGGAGAATATGCTTATGTTTGGGGGACGAATGTTAACATTCCAGATGTGCTTAGGGCGA 720

Qy        721 TTCGTCGATTTCTCCACAATTATCGTTCGAGTGCTCATGATCTTAATTCCAAGTACATCC 780
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        721 TTCGTCGATTTCTCCACAATTATCGTTCGAGTGCTCATGATCTTAATTCCAAGTACATCC 780

Qy        781 AGATCATAGAGGAGACTGTGGAGCGTGAGGAGGATACTCTAAATATCGACATGTCAGACA 840
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        781 AGATCATAGAGGAGACTGTGGAGCGTGAGGAGGATACTCTAAATATCGACATGTCAGACA 840

Qy        841 TTTATGACCATGATCCTGATCTATACGCAAAAATTGTTCGATACCCACTCGACATCATCC 900
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        841 TTTATGACCATGATCCTGATCTATACGCAAAAATTGTTCGATACCCACTCGACATCATCC 900

Qy        901 CCCTGTTGGACACTGAGTGTCAGGAAGTTGCTACCTCTTTACTACCAACGTTTGAGAAGC 960
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        901 CCCTGTTGGACACTGAGTGTCAGGAAGTTGCTACCTCTTTACTACCAACGTTTGAGAAGC 960

Qy        961 ATATTGAGGCCAGACCTTTCAATCTCAAAGCATCGGTGCACATGCGTGAACTCAACCCTT 1020
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        961 ATATTGAGGCCAGACCTTTCAATCTCAAAGCATCGGTGCACATGCGTGAACTCAACCCTT 1020

Qy       1021 CAGATATAGACAAATTGGTTTCTGTTAAAGGAATGGTTATCCGGTGCAGTTCTATCATAC 1080
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1021 CAGATATAGACAAATTGGTTTCTGTTAAAGGAATGGTTATCCGGTGCAGTTCTATCATAC 1080

Qy       1081 CTGAAATTAAGGGGGCCTTCTTCAAATGTTTAGTGTGTGGTCACTCGCCTCCGCTAGTTA 1140
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1081 CTGAAATTAAGGGGGCCTTCTTCAAATGTTTAGTGTGTGGTCACTCGCCTCCGCTAGTTA 1140

Qy       1141 CAGTTGTTAAAGGGCGGGTTGAGGAGCCAACAAGGTGTGAAAAGCCAGAATGTGCAGCAC 1200
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1141 CAGTTGTTAAAGGGCGGGTTGAGGAGCCAACAAGGTGTGAAAAGCCAGAATGTGCAGCAC 1200

Qy       1201 GGAATGCTATGTCTCTTATTCACAATCGATGCACTTTTGCAAATAAGCAGATAGTGCGTC 1260
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1201 GGAATGCTATGTCTCTTATTCACAATCGATGCACTTTTGCAAATAAGCAGATAGTGCGTC 1260

Qy       1261 TTCAAGAAACTCCAGATGCCATTCCTGAAGGAGAGACTCCACACACAGTCAGCATGTGTT 1320
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1261 TTCAAGAAACTCCAGATGCCATTCCTGAAGGAGAGACTCCACACACAGTCAGCATGTGTT 1320

Qy       1321 TATACAACACTATGGTTGATGCTGTGAAGCCTGGAGATCGTATTGAGGTAACAGGAGTTT 1380
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1321 TATACAACACTATGGTTGATGCTGTGAAGCCTGGAGATCGTATTGAGGTAACAGGAGTTT 1380

Qy       1381 TCAAGGCCATGGCAGTTCGAGTTGGTCATGGCGCGCCGCACTCACGTGAGGAATTGCACC 1440
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1381 TCAAGGCCATGGCAGTTCGAGTTGGTCATGGCGCGCCGCACTCACGTGAGGAATTGCACC 1440

Qy       1441 TCCTTGTTCTGCGACGGTTCCATTCTTTTTGGTTTTTAGTTTGCAAATCTTGATCGTGGA 1500
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1441 TCCTTGTTCTGCGACGGTTCCATTCTTTTTGGTTTTTAGTTTGCAAATCTTGATCGTGGA 1500

Qy       1501 GTTGAGAAAAAGGGCGGTTCGTTGTCTTGAGGTGTTCTTGTTGATTGTTCGTCATGGAAA 1560
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1501 GTTGAGAAAAAGGGCGGTTCGTTGTCTTGAGGTGTTCTTGTTGATTGTTCGTCATGGAAA 1560

Qy       1561 ATAATGATGCACTTGACATTGGAGCCGTGTCGTCCCCATATCCTTCGCAATCTGAAGGAG 1620
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1561 ATAATGATGCACTTGACATTGGAGCCGTGTCGTCCCCATATCCTTCGCAATCTGAAGGAG 1620

Qy       1621 TGTCTACGCCATTGCCGCAAGTAACATCACCGAGCTTCGACAATGCAGCCTCACCCGTGG 1680
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1621 TGTCTACGCCATTGCCGCAAGTAACATCACCGAGCTTCGACAATGCAGCCTCACCCGTGG 1680

Qy       1681 CCGGGCGGAGGGCCGTACGGCAGACCCCTACATCTGCAGTTCGAAGGAGAGGGAGAGAAA 1740
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1681 CCGGGCGGAGGGCCGTACGGCAGACCCCTACATCTGCAGTTCGAAGGAGAGGGAGAGAAA 1740

Qy       1741 CGGATTCCGCTCGTCGTAGGAGGAGTCGATCTCGCAGTTTAGGCAATTCTGTTTATAGTT 1800
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1741 CGGATTCCGCTCGTCGTAGGAGGAGTCGATCTCGCAGTTTAGGCAATTCTGTTTATAGTT 1800

Qy       1801 CCCCTTACGATGCGGGGACTCCTGGAACTCCTGGAACTCCAGTGGCTACTCCGGTTTACG 1860
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1801 CCCCTTACGATGCGGGGACTCCTGGAACTCCTGGAACTCCAGTGGCTACTCCGGTTTACG 1860

Qy       1861 CTACCCCAGTCGGTACACCTATGGGTACCCCATCGTTCCATCGTGGCACGCCACAGTACA 1920
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1861 CTACCCCAGTCGGTACACCTATGGGTACCCCATCGTTCCATCGTGGCACGCCACAGTACA 1920

Qy       1921 AACAGCGCAGTGAGCTTGGTTCCCAGGGGAAGCCTCTACATCGGAGACGTCGATCTCAAT 1980
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1921 AACAGCGCAGTGAGCTTGGTTCCCAGGGGAAGCCTCTACATCGGAGACGTCGATCTCAAT 1980

Qy       1981 CCAGAGAACCCGGGCATCGATCTCCTTCAAGGGAACCTAGTGCTGATGGGCGTCCCTCTG 2040
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1981 CCAGAGAACCCGGGCATCGATCTCCTTCAAGGGAACCTAGTGCTGATGGGCGTCCCTCTG 2040

Qy       2041 AATCTGCTGAGCCAGATGACACTTTGGGTGGAGAATATGCTTATGTTTGGGGGACGAATG 2100
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       2041 AATCTGCTGAGCCAGATGACACTTTGGGTGGAGAATATGCTTATGTTTGGGGGACGAATG 2100

Qy       2101 TTAACATTCCAGATGTGCTTAGGGCGATTCGTCGATTTCTCCACAATTATCGTTCGAGTG 2160
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       2101 TTAACATTCCAGATGTGCTTAGGGCGATTCGTCGATTTCTCCACAATTATCGTTCGAGTG 2160

Qy       2161 CTCATGATCTTAATTCCAAGTACATCCAGATCATAGAGGAGACTGTGGAGCGTGAGGAGG 2220
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       2161 CTCATGATCTTAATTCCAAGTACATCCAGATCATAGAGGAGACTGTGGAGCGTGAGGAGG 2220

Qy       2221 ATACTCTAAATATCGACATGTCAGACATTTATGACCATGATCCTGATCTATACGCAAAAA 2280
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       2221 ATACTCTAAATATCGACATGTCAGACATTTATGACCATGATCCTGATCTATACGCAAAAA 2280

Qy       2281 TTGTTCGATACCCACTCGACATCATCCCCCTGTTGGACACTGAGTGTCAGGAAGTTGCTA 2340
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       2281 TTGTTCGATACCCACTCGACATCATCCCCCTGTTGGACACTGAGTGTCAGGAAGTTGCTA 2340

Qy       2341 CCTCTTTACTACCAACGTTTGAGAAGCATATTGAGGCCAGACCTTTCAATCTCAAAGCAT 2400
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       2341 CCTCTTTACTACCAACGTTTGAGAAGCATATTGAGGCCAGACCTTTCAATCTCAAAGCAT 2400

Qy       2401 CGGTGCACATGCGTGAACTCAACCCTTCAGATATAGACAAATTGGTTTCTGTTAAAGGAA 2460
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       2401 CGGTGCACATGCGTGAACTCAACCCTTCAGATATAGACAAATTGGTTTCTGTTAAAGGAA 2460
```

```
Qy       2461 TGGTTATCCGGTGCAGTTCTATCATACCTGAAATTAAGGGGGCCTTCTTCAAATGTTTAG 2520
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       2461 TGGTTATCCGGTGCAGTTCTATCATACCTGAAATTAAGGGGGCCTTCTTCAAATGTTTAG 2520

Qy       2521 TGTGTGGTCACTCGCCTCCGCTAGTTACAGTTGTTAAAGGGCGGGTTGAGGAGCCAACAA 2580
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       2521 TGTGTGGTCACTCGCCTCCGCTAGTTACAGTTGTTAAAGGGCGGGTTGAGGAGCCAACAA 2580

Qy       2581 GGTGTGAAAAGCCAGAATGTGCAGCACGGAATGCTATGTCTCTTATTCACAATCGATGCA 2640
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       2581 GGTGTGAAAAGCCAGAATGTGCAGCACGGAATGCTATGTCTCTTATTCACAATCGATGCA 2640

Qy       2641 CTTTTGCAAATAAGCAGATAGTGCGTCTTCAAGAAACTCCAGATGCCATTCCTGAAGGAG 2700
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       2641 CTTTTGCAAATAAGCAGATAGTGCGTCTTCAAGAAACTCCAGATGCCATTCCTGAAGGAG 2700

Qy       2701 AGACTCCACACACAGTCAGCATGTGTTTATACAACACTATGGTTGATGCTGTGAAGCCTG 2760
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       2701 AGACTCCACACACAGTCAGCATGTGTTTATACAACACTATGGTTGATGCTGTGAAGCCTG 2760

Qy       2761 GAGATCGTATTGAGGTAACAGGAGTTTTCAAGGCCATGGCAGTTCGAGTTGGTCCGAATC 2820
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       2761 GAGATCGTATTGAGGTAACAGGAGTTTTCAAGGCCATGGCAGTTCGAGTTGGTCCGAATC 2820

Qy       2821 AACGAACATTACGAGCATTGTATAAGACCTACATCGATTGCGTGCACGTCAAGAAGTCTG 2880
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       2821 AACGAACATTACGAGCATTGTATAAGACCTACATCGATTGCGTGCACGTCAAGAAGTCTG 2880

Qy       2881 ACAGGGGTCGACTGCAAACTGAAGATCCTATGGAGATGGATAAGGAGAATGATATGTATG 2940
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       2881 ACAGGGGTCGACTGCAAACTGAAGATCCTATGGAGATGGATAAGGAGAATGATATGTATG 2940

Qy       2941 CTGGGTATCATGAAAGTGATACTTCAGAAGCTGCTAATGAAGCAAAGATTCAAAAACTTA 3000
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       2941 CTGGGTATCATGAAAGTGATACTTCAGAAGCTGCTAATGAAGCAAAGATTCAAAAACTTA 3000

Qy       3001 AAGAGCTGTCCAAGCTCCCGGGCATTTATGATAGACTTTCAAGGTCGCTGGCTCCAAGCA 3060
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       3001 AAGAGCTGTCCAAGCTCCCGGGCATTTATGATAGACTTTCAAGGTCGCTGGCTCCAAGCA 3060

Qy       3061 TTTGGGAGCTTGAAGATATTAAAAAGGGTCTTCTTTGCCAGCTCTTTGGTGGGAAGGCTA 3120
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       3061 TTTGGGAGCTTGAAGATATTAAAAAGGGTCTTCTTTGCCAGCTCTTTGGTGGGAAGGCTA 3120

Qy       3121 AGAAAATTCCATCTGGAGCATCTTTCCGAGGTGACATCAATGTTTTACTTGTTGGGGACC 3180
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       3121 AGAAAATTCCATCTGGAGCATCTTTCCGAGGTGACATCAATGTTTTACTTGTTGGGGACC 3180

Qy       3181 CTGGTACCAGTAAATCTCAGCTGCTTCAGTATGTGCACAAGATAGCTCCTCGTGGAATCT 3240
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       3181 CTGGTACCAGTAAATCTCAGCTGCTTCAGTATGTGCACAAGATAGCTCCTCGTGGAATCT 3240

Qy       3241 ACACTAGTGGGCGAGGAAGTTCGGCGGTTGGGCTGACAGCGTATGTAACGAAGGATCCAG 3300
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       3241 ACACTAGTGGGCGAGGAAGTTCGGCGGTTGGGCTGACAGCGTATGTAACGAAGGATCCAG 3300

Qy       3301 AAACTCGAGAGACGGTATTGGAGAGCGGAGCTTTGGTTCTTAGTGATCGTGGGATATGCT 3360
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       3301 AAACTCGAGAGACGGTATTGGAGAGCGGAGCTTTGGTTCTTAGTGATCGTGGGATATGCT 3360

Qy       3361 GTATCGATGAGTTCGACAAAATGTCTGATAATGCCCGAAGCATGCTTCATGAGGTAATGG 3420
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       3361 GTATCGATGAGTTCGACAAAATGTCTGATAATGCCCGAAGCATGCTTCATGAGGTAATGG 3420

Qy       3421 AGCAACAAACGGTATCTGTAGCCAAAGGGGGTATCATTGCCTCGCTGAACGCTCGGACGT 3480
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       3421 AGCAACAAACGGTATCTGTAGCCAAAGGGGGTATCATTGCCTCGCTGAACGCTCGGACGT 3480

Qy       3481 CTGTCCTTGCATGTGCAAATCCTAGTGGGTCCCGATACAATGCGCGCCTTTCTGTGATTG 3540
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       3481 CTGTCCTTGCATGTGCAAATCCTAGTGGGTCCCGATACAATGCGCGCCTTTCTGTGATTG 3540

Qy       3541 ATAACATCCAGCTTCCTCCAACTCTACTTTCTAGATTTGATTTAATTTACTTAATGCTCG 3600
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       3541 ATAACATCCAGCTTCCTCCAACTCTACTTTCTAGATTTGATTTAATTTACTTAATGCTCG 3600

Qy       3601 ACAAACCAGACGAGCAAAACGATCGTCGTCTCGCCAGGCATCTCGTGGCTTTACACTATG 3660
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       3601 ACAAACCAGACGAGCAAAACGATCGTCGTCTCGCCAGGCATCTCGTGGCTTTACACTATG 3660

Qy       3661 AAAACTATGAAGTTTCAAAGCAGGACGCCTTAGATCTACAAACACTTACCGCGTATATCA 3720
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       3661 AAAACTATGAAGTTTCAAAGCAGGACGCCTTAGATCTACAAACACTTACCGCGTATATCA 3720

Qy       3721 CCTATGCTCGTCAGCATGTACATCCTACATTAAGTGATGAAGCTGCTGAAGATTTGATTA 3780
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       3721 CCTATGCTCGTCAGCATGTACATCCTACATTAAGTGATGAAGCTGCTGAAGATTTGATTA 3780

Qy       3781 ATGGCTATGTTGAGATGCGCCAAAAGGGCAACTTTCCTGGAAGCAGTAAAAAGGTGATAA 3840
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       3781 ATGGCTATGTTGAGATGCGCCAAAAGGGCAACTTTCCTGGAAGCAGTAAAAAGGTGATAA 3840

Qy       3841 CAGCCACACCTCGGCAACTCGAAAGTATGATTCGTATCAGTGAAGCCCTAGCTCGAATGA 3900
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       3841 CAGCCACACCTCGGCAACTCGAAAGTATGATTCGTATCAGTGAAGCCCTAGCTCGAATGA 3900

Qy       3901 GATTTTCTGAAGTGGTAGAGAAAGTTGATGCAGCAGAAGCTGTGCGCCTTTTAGACGTCG 3960
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       3901 GATTTTCTGAAGTGGTAGAGAAAGTTGATGCAGCAGAAGCTGTGCGCCTTTTAGACGTCG 3960

Qy       3961 CTTTGCAGCAATCTGCTACTGATCATGCAACAGGTACGATAGACATGGATCTTATCACGA 4020
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       3961 CTTTGCAGCAATCTGCTACTGATCATGCAACAGGTACGATAGACATGGATCTTATCACGA 4020

Qy       4021 CTGGAGTGTCGGCCAGCGAGCGTATTCGTCGGGCCAACTTGCTAGCTGCTCTGCGAGAGC 4080
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       4021 CTGGAGTGTCGGCCAGCGAGCGTATTCGTCGGGCCAACTTGCTAGCTGCTCTGCGAGAGC 4080

Qy       4081 TTATAGCAGATAAAATTTCACCTGGCAGCTCCTCTGGCTTGAAGACCAGTCAGCTTCTTG 4140
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       4081 TTATAGCAGATAAAATTTCACCTGGCAGCTCCTCTGGCTTGAAGACCAGTCAGCTTCTTG 4140

Qy       4141 AGGATATCCGGAGCCAAAGCAGTGTGGACGTTAGTTTGCAGGATATTAAAAATGCTCTGG 4200
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       4141 AGGATATCCGGAGCCAAAGCAGTGTGGACGTTAGTTTGCAGGATATTAAAAATGCTCTGG 4200

Qy       4201 GTAGCCTCCAAGGAGAAGGCTTTCTTACTGTCCATGGTGACATAGTCAAGAGAGTTTGAG 4260
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       4201 GTAGCCTCCAAGGAGAAGGCTTTCTTACTGTCCATGGTGACATAGTCAAGAGAGTTTGAG 4260

Qy       4261 ACAGTTTCTAACTGTTCGAATCCATGAGCTATAACTCTGAACGAAAGGAAAACCTCCAG 4320
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       4261 ACAGTTTCTAACTGTTCGAATCCATGAGCTATAACTCTGAACGAAAGGAAAACCTCCAG 4320

Qy       4321 TTTCCCATGCGCAATTCCCAGAGCTCGC 4348
              ||||||||||||||||||||||||||||
Db       4321 TTTCCCATGCGCAATTCCCAGAGCTCGC 4348
```

```
RESULT 2
US-10-768-511-5
; Sequence 5, Application US/10768511
; Publication No. US20040128721A1
; GENERAL INFORMATION:
;  APPLICANT: COSTA E SILVA, OSWALDO DA
;  APPLICANT:  BOHNERT, HANS J.
;  APPLICANT:  VAN THIELEN, NOCHA
;  APPLICANT:  CHEN, ROUYING
;  APPLICANT:  SARRIA-MILLAN, RODRIGO
;  TITLE OF INVENTION: CELL CYCLE STRESS-RELATED PROTEINS AND METHODS OF USE
;  TITLE OF INVENTION:  IN PLANTS
;  FILE REFERENCE: 16313-0031
;  CURRENT APPLICATION NUMBER: US/10/768,511
;  CURRENT FILING DATE:  2004-01-30
;  PRIOR APPLICATION NUMBER: 60/196,001
```

```
;  PRIOR FILING DATE: 2000-04-07
;  NUMBER OF SEQ ID NOS: 34
;  SOFTWARE: PatentIn Ver. 2.1
; SEQ ID NO 5
;   LENGTH: 4348
;   TYPE: DNA
;   ORGANISM: Physcomitrella patens
US-10-768-511-5

  Query Match             100.0%;  Score 4348;  DB 7;  Length 4348;
  Best Local Similarity   100.0%;  Pred. No. 0;
  Matches 4348;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy          1 ATGGCGCGCCGCACTCACGTGAGGAATTGCACCTCCTTGTTCTGCGACGGTTCCATTCTT 60
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db          1 ATGGCGCGCCGCACTCACGTGAGGAATTGCACCTCCTTGTTCTGCGACGGTTCCATTCTT 60

Qy         61 TTTGGTTTTTAGTTTGCAAATCTTGATCGTGGAGTTGAGAAAAAGGGCGGTTCGTTGTCT 120
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db         61 TTTGGTTTTTAGTTTGCAAATCTTGATCGTGGAGTTGAGAAAAAGGGCGGTTCGTTGTCT 120

Qy        121 TGAGGTGTTCTTGTTGATTGTTCGTCATGGAAAATAATGATGCACTTGACATTGGAGCCG 180
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        121 TGAGGTGTTCTTGTTGATTGTTCGTCATGGAAAATAATGATGCACTTGACATTGGAGCCG 180

Qy        181 TGTCGTCCCCATATCCTTCGCAATCTGAAGGAGTGTCTACGCCATTGCCGCAAGTAACAT 240
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        181 TGTCGTCCCCATATCCTTCGCAATCTGAAGGAGTGTCTACGCCATTGCCGCAAGTAACAT 240

Qy        241 CACCGAGCTTCGACAATGCAGCCTCACCCGTGGCCGGGCGGAGGGCCGTACGGCAGACCC 300
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        241 CACCGAGCTTCGACAATGCAGCCTCACCCGTGGCCGGGCGGAGGGCCGTACGGCAGACCC 300

Qy        301 CTACATCTGCAGTTCGAAGGAGAGGGAGAGAAACGGATTCCGCTCGTCGTAGGAGGAGTC 360
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        301 CTACATCTGCAGTTCGAAGGAGAGGGAGAGAAACGGATTCCGCTCGTCGTAGGAGGAGTC 360

Qy        361 GATCTCGCAGTTTAGGCAATTCTGTTTATAGTTCCCCTTACGATGCGGGGACTCCTGGAA 420
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        361 GATCTCGCAGTTTAGGCAATTCTGTTTATAGTTCCCCTTACGATGCGGGGACTCCTGGAA 420

Qy        421 CTCCTGGAACTCCAGTGGCTACTCCGGTTTACGCTACCCCAGTCGGTACACCTATGGGTA 480
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        421 CTCCTGGAACTCCAGTGGCTACTCCGGTTTACGCTACCCCAGTCGGTACACCTATGGGTA 480

Qy        481 CCCCATCGTTCCATCGTGGCACGCCACAGTACAAACAGCGCAGTGAGCTTGGTTCCCAGG 540
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        481 CCCCATCGTTCCATCGTGGCACGCCACAGTACAAACAGCGCAGTGAGCTTGGTTCCCAGG 540

Qy        541 GGAAGCCTCTACATCGGAGACGTCGATCTCAATCCAGAGAACCCGGGCATCGATCTCCTT 600
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        541 GGAAGCCTCTACATCGGAGACGTCGATCTCAATCCAGAGAACCCGGGCATCGATCTCCTT 600

Qy        601 CAAGGGAACCTAGTGCTGATGGGCGTCCCTCTGAATCTGCTGAGCCAGATGACACTTTGG 660
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        601 CAAGGGAACCTAGTGCTGATGGGCGTCCCTCTGAATCTGCTGAGCCAGATGACACTTTGG 660

Qy        661 GTGGAGAATATGCTTATGTTTGGGGGACGAATGTTAACATTCCAGATGTGCTTAGGGCGA 720
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        661 GTGGAGAATATGCTTATGTTTGGGGGACGAATGTTAACATTCCAGATGTGCTTAGGGCGA 720

Qy        721 TTCGTCGATTTCTCCACAATTATCGTTCGAGTGCTCATGATCTTAATTCCAAGTACATCC 780
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        721 TTCGTCGATTTCTCCACAATTATCGTTCGAGTGCTCATGATCTTAATTCCAAGTACATCC 780

Qy        781 AGATCATAGAGGAGACTGTGGAGCGTGAGGAGGATACTCTAAATATCGACATGTCAGACA 840
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        781 AGATCATAGAGGAGACTGTGGAGCGTGAGGAGGATACTCTAAATATCGACATGTCAGACA 840

Qy        841 TTTATGACCATGATCCTGATCTATACGCAAAAATTGTTCGATACCCACTCGACATCATCC 900
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        841 TTTATGACCATGATCCTGATCTATACGCAAAAATTGTTCGATACCCACTCGACATCATCC 900

Qy        901 CCCTGTTGGACACTGAGTGTCAGGAAGTTGCTACCTCTTTACTACCAACGTTTGAGAAGC 960
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        901 CCCTGTTGGACACTGAGTGTCAGGAAGTTGCTACCTCTTTACTACCAACGTTTGAGAAGC 960

Qy        961 ATATTGAGGCCAGACCTTTCAATCTCAAAGCATCGGTGCACATGCGTGAACTCAACCCTT 1020
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        961 ATATTGAGGCCAGACCTTTCAATCTCAAAGCATCGGTGCACATGCGTGAACTCAACCCTT 1020

Qy       1021 CAGATATAGACAAATTGGTTTCTGTTAAAGGAATGGTTATCCGGTGCAGTTCTATCATAC 1080
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1021 CAGATATAGACAAATTGGTTTCTGTTAAAGGAATGGTTATCCGGTGCAGTTCTATCATAC 1080

Qy       1081 CTGAAATTAAGGGGGCCTTCTTCAAATGTTTAGTGTGTGGTCACTCGCCTCCGCTAGTTA 1140
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1081 CTGAAATTAAGGGGGCCTTCTTCAAATGTTTAGTGTGTGGTCACTCGCCTCCGCTAGTTA 1140

Qy       1141 CAGTTGTTAAAGGGCGGGTTGAGGAGCCAACAAGGTGTGAAAAGCCAGAATGTGCAGCAC 1200
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1141 CAGTTGTTAAAGGGCGGGTTGAGGAGCCAACAAGGTGTGAAAAGCCAGAATGTGCAGCAC 1200

Qy       1201 GGAATGCTATGTCTCTTATTCACAATCGATGCACTTTTGCAAATAAGCAGATAGTGCGTC 1260
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1201 GGAATGCTATGTCTCTTATTCACAATCGATGCACTTTTGCAAATAAGCAGATAGTGCGTC 1260

Qy       1261 TTCAAGAAACTCCAGATGCCATTCCTGAAGGAGAGACTCCACACACAGTCAGCATGTGTT 1320
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1261 TTCAAGAAACTCCAGATGCCATTCCTGAAGGAGAGACTCCACACACAGTCAGCATGTGTT 1320

Qy       1321 TATACAACACTATGGTTGATGCTGTGAAGCCTGGAGATCGTATTGAGGTAACAGGAGTTT 1380
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1321 TATACAACACTATGGTTGATGCTGTGAAGCCTGGAGATCGTATTGAGGTAACAGGAGTTT 1380

Qy       1381 TCAAGGCCATGGCAGTTCGAGTTGGTCATGGCGCGCCGCACTCACGTGAGGAATTGCACC 1440
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1381 TCAAGGCCATGGCAGTTCGAGTTGGTCATGGCGCGCCGCACTCACGTGAGGAATTGCACC 1440

Qy       1441 TCCTTGTTCTGCGACGGTTCCATTCTTTTTGGTTTTTAGTTTGCAAATCTTGATCGTGGA 1500
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1441 TCCTTGTTCTGCGACGGTTCCATTCTTTTTGGTTTTTAGTTTGCAAATCTTGATCGTGGA 1500

Qy       1501 GTTGAGAAAAAGGGCGGTTCGTTGTCTTGAGGTGTTCTTGTTGATTGTTCGTCATGGAAA 1560
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1501 GTTGAGAAAAAGGGCGGTTCGTTGTCTTGAGGTGTTCTTGTTGATTGTTCGTCATGGAAA 1560

Qy       1561 ATAATGATGCACTTGACATTGGAGCCGTGTCGTCCCCATATCCTTCGCAATCTGAAGGAG 1620
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1561 ATAATGATGCACTTGACATTGGAGCCGTGTCGTCCCCATATCCTTCGCAATCTGAAGGAG 1620

Qy       1621 TGTCTACGCCATTGCCGCAAGTAACATCACCGAGCTTCGACAATGCAGCCTCACCCGTGG 1680
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1621 TGTCTACGCCATTGCCGCAAGTAACATCACCGAGCTTCGACAATGCAGCCTCACCCGTGG 1680

Qy       1681 CCGGGCGGAGGGCCGTACGGCAGACCCCTACATCTGCAGTTCGAAGGAGAGGGAGAGAAA 1740
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1681 CCGGGCGGAGGGCCGTACGGCAGACCCCTACATCTGCAGTTCGAAGGAGAGGGAGAGAAA 1740

Qy       1741 CGGATTCCGCTCGTCGTAGGAGGAGTCGATCTCGCAGTTTAGGCAATTCTGTTTATAGTT 1800
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1741 CGGATTCCGCTCGTCGTAGGAGGAGTCGATCTCGCAGTTTAGGCAATTCTGTTTATAGTT 1800

Qy       1801 CCCCTTACGATGCGGGGACTCCTGGAACTCCTGGAACTCCAGTGGCTACTCCGGTTTACG 1860
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1801 CCCCTTACGATGCGGGGACTCCTGGAACTCCTGGAACTCCAGTGGCTACTCCGGTTTACG 1860

Qy       1861 CTACCCCAGTCGGTACACCTATGGGTACCCCATCGTTCCATCGTGGCACGCCACAGTACA 1920
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1861 CTACCCCAGTCGGTACACCTATGGGTACCCCATCGTTCCATCGTGGCACGCCACAGTACA 1920

Qy       1921 AACAGCGCAGTGAGCTTGGTTCCCAGGGGAAGCCTCTACATCGGAGACGTCGATCTCAAT 1980
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1921 AACAGCGCAGTGAGCTTGGTTCCCAGGGGAAGCCTCTACATCGGAGACGTCGATCTCAAT 1980

Qy       1981 CCAGAGAACCCGGGCATCGATCTCCTTCAAGGGAACCTAGTGCTGATGGGCGTCCCTCTG 2040
```

```
                 ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        1981 CCAGAGAACCCGGGCATCGATCTCCTTCAAGGGAACCTAGTGCTGATGGGCGTCCCTCTG 2040

Qy        2041 AATCTGCTGAGCCAGATGACACTTTGGGTGGAGAATATGCTTATGTTTGGGGGACGAATG 2100
               ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        2041 AATCTGCTGAGCCAGATGACACTTTGGGTGGAGAATATGCTTATGTTTGGGGGACGAATG 2100

Qy        2101 TTAACATTCCAGATGTGCTTAGGGCGATTCGTCGATTTCTCCACAATTATCGTTCGAGTG 2160
               ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        2101 TTAACATTCCAGATGTGCTTAGGGCGATTCGTCGATTTCTCCACAATTATCGTTCGAGTG 2160

Qy        2161 CTCATGATCTTAATTCCAAGTACATCCAGATCATAGAGGAGACTGTGGAGCGTGAGGAGG 2220
               ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        2161 CTCATGATCTTAATTCCAAGTACATCCAGATCATAGAGGAGACTGTGGAGCGTGAGGAGG 2220

Qy        2221 ATACTCTAAATATCGACATGTCAGACATTTATGACCATGATCCTGATCTATACGCAAAAA 2280
               ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        2221 ATACTCTAAATATCGACATGTCAGACATTTATGACCATGATCCTGATCTATACGCAAAAA 2280

Qy        2281 TTGTTCGATACCCACTCGACATCATCCCCCTGTTGGACACTGAGTGTCAGGAAGTTGCTA 2340
               ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        2281 TTGTTCGATACCCACTCGACATCATCCCCCTGTTGGACACTGAGTGTCAGGAAGTTGCTA 2340

Qy        2341 CCTCTTTACTACCAACGTTTGAGAAGCATATTGAGGCCAGACCTTTCAATCTCAAAGCAT 2400
               ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        2341 CCTCTTTACTACCAACGTTTGAGAAGCATATTGAGGCCAGACCTTTCAATCTCAAAGCAT 2400

Qy        2401 CGGTGCACATGCGTGAACTCAACCCTTCAGATATAGACAAATTGGTTTCTGTTAAAGGAA 2460
               ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        2401 CGGTGCACATGCGTGAACTCAACCCTTCAGATATAGACAAATTGGTTTCTGTTAAAGGAA 2460

Qy        2461 TGGTTATCCGGTGCAGTTCTATCATACCTGAAATTAAGGGGGCCTTCTTCAAATGTTTAG 2520
               ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        2461 TGGTTATCCGGTGCAGTTCTATCATACCTGAAATTAAGGGGGCCTTCTTCAAATGTTTAG 2520

Qy        2521 TGTGTGGTCACTCGCCTCCGCTAGTTACAGTTGTTAAAGGGCGGGTTGAGGAGCCAACAA 2580
               ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        2521 TGTGTGGTCACTCGCCTCCGCTAGTTACAGTTGTTAAAGGGCGGGTTGAGGAGCCAACAA 2580

Qy        2581 GGTGTGAAAAGCCAGAATGTGCAGCACGGAATGCTATGTCTCTTATTCACAATCGATGCA 2640
               ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        2581 GGTGTGAAAAGCCAGAATGTGCAGCACGGAATGCTATGTCTCTTATTCACAATCGATGCA 2640

Qy        2641 CTTTTGCAAATAAGCAGATAGTGCGTCTTCAAGAAACTCCAGATGCCATTCCTGAAGGAG 2700
               ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        2641 CTTTTGCAAATAAGCAGATAGTGCGTCTTCAAGAAACTCCAGATGCCATTCCTGAAGGAG 2700

Qy        2701 AGACTCCACACACAGTCAGCATGTGTTTATACAACACTATGGTTGATGCTGTGAAGCCTG 2760
               ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        2701 AGACTCCACACACAGTCAGCATGTGTTTATACAACACTATGGTTGATGCTGTGAAGCCTG 2760

Qy        2761 GAGATCGTATTGAGGTAACAGGAGTTTTCAAGGCCATGGCAGTTCGAGTTGGTCCGAATC 2820
               ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        2761 GAGATCGTATTGAGGTAACAGGAGTTTTCAAGGCCATGGCAGTTCGAGTTGGTCCGAATC 2820

Qy        2821 AACGAACATTACGAGCATTGTATAAGACCTACATCGATTGCGTGCACGTCAAGAAGTCTG 2880
               ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        2821 AACGAACATTACGAGCATTGTATAAGACCTACATCGATTGCGTGCACGTCAAGAAGTCTG 2880

Qy        2881 ACAGGGGTCGACTGCAAACTGAAGATCCTATGGAGATGGATAAGGAGAATGATATGTATG 2940
               ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        2881 ACAGGGGTCGACTGCAAACTGAAGATCCTATGGAGATGGATAAGGAGAATGATATGTATG 2940

Qy        2941 CTGGGTATCATGAAAGTGATACTTCAGAAGCTGCTAATGAAGCAAAGATTCAAAAACTTA 3000
               ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        2941 CTGGGTATCATGAAAGTGATACTTCAGAAGCTGCTAATGAAGCAAAGATTCAAAAACTTA 3000

Qy        3001 AAGAGCTGTCCAAGCTCCCGGGCATTTATGATAGACTTTCAAGGTCGCTGGCTCCAAGCA 3060
               ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        3001 AAGAGCTGTCCAAGCTCCCGGGCATTTATGATAGACTTTCAAGGTCGCTGGCTCCAAGCA 3060

Qy        3061 TTTGGGAGCTTGAAGATATTAAAAAGGGTCTTCTTTGCCAGCTCTTTGGTGGGAAGGCTA 3120
               ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        3061 TTTGGGAGCTTGAAGATATTAAAAAGGGTCTTCTTTGCCAGCTCTTTGGTGGGAAGGCTA 3120

Qy        3121 AGAAAATTCCATCTGGAGCATCTTTCCGAGGTGACATCAATGTTTTACTTGTTGGGGACC 3180
               ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        3121 AGAAAATTCCATCTGGAGCATCTTTCCGAGGTGACATCAATGTTTTACTTGTTGGGGACC 3180

Qy        3181 CTGGTACCAGTAAATCTCAGCTGCTTCAGTATGTGCACAAGATAGCTCCTCGTGGAATCT 3240
               ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        3181 CTGGTACCAGTAAATCTCAGCTGCTTCAGTATGTGCACAAGATAGCTCCTCGTGGAATCT 3240

Qy        3241 ACACTAGTGGGCGAGGAAGTTCGGCGGTTGGGCTGACAGCGTATGTAACGAAGGATCCAG 3300
               ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        3241 ACACTAGTGGGCGAGGAAGTTCGGCGGTTGGGCTGACAGCGTATGTAACGAAGGATCCAG 3300

Qy        3301 AAACTCGAGAGACGGTATTGGAGAGCGGAGCTTTGGTTCTTAGTGATCGTGGGATATGCT 3360
               ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        3301 AAACTCGAGAGACGGTATTGGAGAGCGGAGCTTTGGTTCTTAGTGATCGTGGGATATGCT 3360

Qy        3361 GTATCGATGAGTTCGACAAAATGTCTGATAATGCCCGAAGCATGCTTCATGAGGTAATGG 3420
               ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        3361 GTATCGATGAGTTCGACAAAATGTCTGATAATGCCCGAAGCATGCTTCATGAGGTAATGG 3420

Qy        3421 AGCAACAAACGGTATCTGTAGCCAAAGGGGGTATCATTGCCTCGCTGAACGCTCGGACGT 3480
               ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        3421 AGCAACAAACGGTATCTGTAGCCAAAGGGGGTATCATTGCCTCGCTGAACGCTCGGACGT 3480

Qy        3481 CTGTCCTTGCATGTGCAAATCCTAGTGGGTCCCGATACAATGCGCGCCTTTCTGTGATTG 3540
               ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        3481 CTGTCCTTGCATGTGCAAATCCTAGTGGGTCCCGATACAATGCGCGCCTTTCTGTGATTG 3540

Qy        3541 ATAACATCCAGCTTCCTCCAACTCTACTTTCTAGATTTGATTTAATTTACTTAATGCTCG 3600
               ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        3541 ATAACATCCAGCTTCCTCCAACTCTACTTTCTAGATTTGATTTAATTTACTTAATGCTCG 3600

Qy        3601 ACAAACCAGACGAGCAAAACGATCGTCGTCTCGCCAGGCATCTCGTGGCTTTACACTATG 3660
               ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        3601 ACAAACCAGACGAGCAAAACGATCGTCGTCTCGCCAGGCATCTCGTGGCTTTACACTATG 3660

Qy        3661 AAAACTATGAAGTTTCAAAGCAGGACGCCTTAGATCTACAAACACTTACCGCGTATATCA 3720
               ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        3661 AAAACTATGAAGTTTCAAAGCAGGACGCCTTAGATCTACAAACACTTACCGCGTATATCA 3720

Qy        3721 CCTATGCTCGTCAGCATGTACATCCTACATTAAGTGATGAAGCTGCTGAAGATTTGATTA 3780
               ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        3721 CCTATGCTCGTCAGCATGTACATCCTACATTAAGTGATGAAGCTGCTGAAGATTTGATTA 3780

Qy        3781 ATGGCTATGTTGAGATGCGCCAAAAGGGCAACTTTCCTGGAAGCAGTAAAAAGGTGATAA 3840
               ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        3781 ATGGCTATGTTGAGATGCGCCAAAAGGGCAACTTTCCTGGAAGCAGTAAAAAGGTGATAA 3840

Qy        3841 CAGCCACACCTCGGCAACTCGAAAGTATGATTCGTATCAGTGAAGCCCTAGCTCGAATGA 3900
               ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        3841 CAGCCACACCTCGGCAACTCGAAAGTATGATTCGTATCAGTGAAGCCCTAGCTCGAATGA 3900

Qy        3901 GATTTTCTGAAGTGGTAGAGAAAGTTGATGCAGCAGAAGCTGTGCGCCTTTTAGACGTCG 3960
               ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        3901 GATTTTCTGAAGTGGTAGAGAAAGTTGATGCAGCAGAAGCTGTGCGCCTTTTAGACGTCG 3960

Qy        3961 CTTTGCAGCAATCTGCTACTGATCATGCAACAGGTACGATAGACATGGATCTTATCACGA 4020
               ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        3961 CTTTGCAGCAATCTGCTACTGATCATGCAACAGGTACGATAGACATGGATCTTATCACGA 4020

Qy        4021 CTGGAGTGTCGGCCAGCGAGCGTATTCGTCGGGCCAACTTGCTAGCTGCTCTGCGAGAGC 4080
               ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        4021 CTGGAGTGTCGGCCAGCGAGCGTATTCGTCGGGCCAACTTGCTAGCTGCTCTGCGAGAGC 4080

Qy        4081 TTATAGCAGATAAAATTTCACCTGGCAGCTCCTCTGGCTTGAAGACCAGTCAGCTTCTTG 4140
               ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        4081 TTATAGCAGATAAAATTTCACCTGGCAGCTCCTCTGGCTTGAAGACCAGTCAGCTTCTTG 4140

Qy        4141 AGGATATCCGGAGCCAAAGCAGTGTGGACGTTAGTTTGCAGGATATTAAAAATGCTCTGG 4200
               ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        4141 AGGATATCCGGAGCCAAAGCAGTGTGGACGTTAGTTTGCAGGATATTAAAAATGCTCTGG 4200
```

```
Qy       4201 GTAGCCTCCAAGGAGAAGGCTTTCTTACTGTCCATGGTGACATAGTCAAGAGAGTTTGAG 4260
Db       4201 GTAGCCTCCAAGGAGAAGGCTTTCTTACTGTCCATGGTGACATAGTCAAGAGAGTTTGAG 4260

Qy       4261 ACAGTTTCTAACTGTTCGAATCCATGAGCTATAACTCTGAACGAAAGGGAAAACCTCCAG 4320
Db       4261 ACAGTTTCTAACTGTTCGAATCCATGAGCTATAACTCTGAACGAAAGGGAAAACCTCCAG 4320

Qy       4321 TTTCCCATGCGCAATTCCCAGAGCTCGC 4348
Db       4321 TTTCCCATGCGCAATTCCCAGAGCTCGC 4348
```

```
RESULT 3
US-10-425-114-28735
; Sequence 28735, Application US/10425114
; Publication No. US20040034888A1
; GENERAL INFORMATION:
;  APPLICANT: Liu, Jingdong
;  APPLICANT:  Zhou, Yihua
;  APPLICANT:  Kovalic, David K.
;  APPLICANT:  Screen, Steven E
;  APPLICANT:  Tabaska, Jack E
;  APPLICANT:  Cao, Yongwei
;  TITLE OF INVENTION: Nucleic Acid Molecules and Other Molecules Associated With
;  TITLE OF INVENTION:  Plants and Uses Thereof for Plant Improvement
;  FILE REFERENCE: 38-21(53313)B
;  CURRENT APPLICATION NUMBER: US/10/425,114
;  CURRENT FILING DATE:  2003-04-28
;  NUMBER OF SEQ ID NOS: 73128
; SEQ ID NO 28735
;   LENGTH: 2841
;   TYPE: DNA
;   ORGANISM: Zea mays
;   FEATURE:
;   OTHER INFORMATION: Clone ID: LIB4763-010-C6_FLI
US-10-425-114-28735

  Query Match            20.2%;  Score 880.2;  DB 7;  Length 2841;
  Best Local Similarity   65.3%;  Pred. No. 2.7e-265;
  Matches 1334;  Conservative    0;  Mismatches  688;  Indels   21;  Gaps    2;

Qy       2217 GAGGATACTCTAAATATCGACATGTCAGACATTTATGACCATGATCCTGATCTATACGCA 2276
Db        625 GGGGAGTCGCTCGACGTCGATGCGCACGACGTGTTTGACCACGACCCAGACCTCTACAGC 684

Qy       2277 AAAATTGTTCGATACCCACTCGACATCATCCCCCTGTTGGACACTGAGTGTCAGGAAGTT 2336
Db        685 AAGATGGTTCGCTATCCGCTCGAGGTGCTCGCCATCTTCGACATCGTGCTCATGGACCTC 744

Qy       2337 GCTACCTCTTTACTACCAACGTTTGAGAAGCATATTGAGGCCAGACCTTTCAATCTCAAA 2396
Db        745 GTCGCGCGCATCGAGCCGCTCTTCGAGAAGCACATCCAGACCAGGATCTACAACCTCAAG 804

Qy       2397 GCATCGGTGCACATGCGTGAACTCAACCCTTCAGATATAGACAAATTGGTTTCTGTTAAA 2456
Db        805 TCGTCCATTTGCTTGAGGAATCTCAACCCATCTGATATTGAGAAGATGGTATCCATCAAG 864

Qy       2457 GGAATGGTTATCCGGTGCAGTTCTATCATACCTGAAATTAAGGGGGCCTTCTTCAAATGT 2516
Db        865 GGTATGATAATTAGATGCAGCTCGGTCATACCGGAGCTCAAGGAGGCTGTGTTCCGCTGC 924

Qy       2517 TTAGTGTGTGGTCACTCGCCTCCGCTAGTTACAGTTGTTAAAGGGCGGGTTGAGGAGCCA 2576
Db        925 CTGGTTTGTGGTTTCTACTCAGAGCCCGTCATGGTTGATAGAGGAAGAGTAACTGAACCA 984

Qy       2577 ACAAGGTGTGAAAAGCCAGAATGTGCAGCACGGAATGCTATGTCTCTTATTCACAATCGA 2636
Db        985 CACATTTGTCAGAAAGAACAATGTAAAGCCACAAATTCTATGACCCTAGTGCACAACAGA 1044

Qy       2637 TGCACTTTTGCAAATAAGCAGATAGTGCGTCTTCAAGAAACTCCAGATGCCATTCCTGAA 2696
Db       1045 TGCAGATTTTCAGACAAGCAGATCATAAAGTTGCAGGAAACACCAGACGAGATACCAGAA 1104

Qy       2697 GGAGAGACTCCACACACAGTCAGCATGTGTTTATACAACACTATGGTTGATGCTGTGAAG 2756
Db       1105 GGTGGCACTCCACATACAGTTAGTGTCTTGATGCATGATAAGCTTGTTGATGCTGGAAAG 1164

Qy       2757 CCTGGAGATCGTATTGAGGTAACAGGAGTTTTCAAGGCCATGGCAGTTCGAGTTGGTCCG 2816
Db       1165 CCTGGAGATAGGGTTGAGATAACTGGAATATACAGAGCTATGAGTATTCGAATTGGACCA 1224

Qy       2817 AATCAACGAACATTACGAGCATTGTATAAGACCTACATCGATTGCGTGCACGTCAAGAAG 2876
Db       1225 ACTCAAAGGACAGTGAAGTCTATATTCAAGACATATATTGATTGCCTTCACATAAAGAAG 1284

Qy       2877 TCTGACAGGGGTCGACTGCAAACTGAAGATCCTATGGAGATGGATAAGGAGAATGATATG 2936
Db       1285 ACAGACAAGTCTAGGCTTCATGTGGAGGACACCATGGATATTGATAATTCTAACGCTAGC 1344

Qy       2937 TATGCTGGGTATCATGAAAGTGATACTTCAGAAGCTGCTAATGAAGCAAAGATTCAAAAA 2996
Db       1345 AAATC------------------TACTGAAGAGGATTTTCTTAGTGATAAGGTTGAGAAA 1386

Qy       2997 CTTAAAGAGCTGTCCAAGCTCCCGGGCATTTATGATAGACTTTCAAGGTCGCTGGCTCCA 3056
Db       1387 CTAAAAGAGCTTTCGAAGTTGCCTGATATCTATGAAAGATTGACTAGATCATTAGCTCCA 1446

Qy       3057 AGCATTTGGGAGCTTGAAGATATTAAAAAGGGTCTTCTTTGCCAGCTCTTTGGTGGGAAG 3116
Db       1447 AACATATGGGAGTTGGATGATGTCAAAAGAGGTCTCCTTTGCCAGCTTTTCGGCGGTAAT 1506

Qy       3117 GCTAAGAAAATTCCATCTGGAGCATCTTTCCGAGGTGACATCAATGTTTTACTTGTTGGG 3176
Db       1507 CCCTTGAAGCTTCCTTCTGGAGCTAGTTTCCGGGGTGACATCAATATTTTACTTGTGGGG 1566

Qy       3177 GACCCTGGTACCAGTAAATCTCAGCTGCTTCAGTATGTGCACAAGATAGCTCCTCGTGGA 3236
Db       1567 GACCCTGGAACAAGTAAATCCCAGCTTCTCCAGTACATGCATAAACTGTCTCCTCGTGGT 1626

Qy       3237 ATCTACACTAGTGGGCGAGGAAGTTCGGCGGTTGGGCTGACAGCGTATGTAACGAAGGAT 3296
Db       1627 ATCTATACGAGTGGTAGAGGAAGTTCTGCTGTTGGTCTTACTGCTTATGTTACCAAAGAC 1686

Qy       3297 CCAGAAACTCGAGAGACGGTATTGGAGAGCGGAGCTTTGGTTCTTAGTGATCGTGGGATA 3356
Db       1687 CCTGAGACTGGCGAAACTGTTCTAGAAAGTGGAGCACTTGTTTTGAGTGACAAAGGTGTT 1746

Qy       3357 TGCTGTATCGATGAGTTCGACAAAATGTCTGATAATGCCCGAAGCATGCTTCATGAGGTA 3416
Db       1747 TGTTGCATAGATGAGTTTGATAAGATGTCTGATAATGCCCGAAGCATGTTACACGAGGTG 1806

Qy       3417 ATGGAGCAACAAACGGTATCTGTAGCCAAAGGGGGTATCATTGCCTCGCTGAACGCTCGG 3476
Db       1807 ATGGAACAGCAGACAGTATCCATTGCGAAGGCTGGAATAATTGCATCTTTAAACGCTAGG 1866

Qy       3477 ACGTCTGTCCTTGCATGTGCAAATCCTAGTGGGTCCCGATACAATGCGCGCCTTTCTGTG 3536
Db       1867 ACATCTGTCCTGGCATGTGCCAATCCTACTGAATCACGTTACAATCCAAGGCTCTCTGTA 1926

Qy       3537 ATTGATAACATCCAGCTTCCTCCAACTCTACTTTCTAGATTTGATTTAATTTACTTAATG 3596
Db       1927 ATTGACAACATCCACTTAGCGCCAACGCTACTTTCAAGATTCGACCTGATTTATCTTATC 1986

Qy       3597 CTCGACAAACCAGACGAGCAAAACGATCGTCGTCTCGCCAGGCATCTCGTGGCTTTACAC 3656
Db       1987 TTGGACAAGGCGGATGAGCAAACTGATAGGCGCCTGGCAAAGCATATTGTTTCGTTGCAT 2046

Qy       3657 TATGAAAACTATGAAGTTTCAAAGCAGGACGCCTTAGATCTACAAACACTTACCGCGTAT 3716
Db       2047 TTTGAGAATCCAAATTTAGAGGAGCTCGAGGTCTTGGACTTGCAGACACTAGTTTCCTAC 2106

Qy       3717 ATCACCTATGCTCGTCAGCATGTACATCCTACATTAAGTGATGAAGCTGCTGAAGATTTG 3776
Db       2107 ATAAGCTATGCAAGGAAGTATATTCAGCCACAGTTATCTGATGAAGCTGCAGAAGAGTTA 2166
```

```
Qy       3777 ATTAATGGCTATGTTGAGATGCGCCAAAAGGGCAACTTTCCTGGAAGCAGTAAAAAGGTG 3836
Db       2167 ACTCGTGGCTATGTCGAGATGAGAAAAAGAGGGAATAGCCCTGGGAGCAGAAAGAAGGTC 2226

Qy       3837 ATAACAGCCACACCTCGGCAACTCGAAAGTATGATTCGTATCAGTGAAGCCCTAGCTCGA 3896
Db       2227 ATAACAGCAACCGCTAGACAAATAGAGAGTTTGATCCGTCTCAGCGAAGCATTAGCCCGA 2286

Qy       3897 ATGAGATTTTCTGAAGTGGTAGAGAAAGTTGATGCAGCAGAAGCTGTGCGCCTTTTAGAC 3956
Db       2287 ATGCGGTTCTCTGAAGTGGTCGAGGTGCGGGATGTTGTGGAGGCATTCAGGCTTCTTGAA 2346

Qy       3957 GTCGCTTTGCAGCAATCTGCTACTGATCATGCAACAGGTACGATAGACATGGATCTTATC 4016
Db       2347 GTCGCCATGCAGCAGTCTGCGACGGATCATGCAACTGGTACGATTGATATGGATCTGATC 2406

Qy       4017 ACGACTGGAGTGTCGGCCAGCGAGCGTATTCGTCGGGCCAACTTGCTAGCTGCTCTGCGA 4076
Db       2407 ATGACGGGGATATCCGCAAGCGAAAGGCAGAGGCGGGAGAACCTCGTTGCCGCAACCCGT 2466

Qy       4077 GAGCTTATAGCAGATAAAATTTCACCTGGCAGCTCCTCTGGCTTGAAGACCAGTCAGCTT 4136
Db       2467 AACCTGATTGCGGAGAAAATGCAGCTTGGAGGCCCCTC---GATGCGCATGATTGAGTTG 2523

Qy       4137 CTTGAGGATATCCGGAGCCAAAGCAGTGTGGACGTTAGTTTGCAGGATATTAAAAATGCT 4196
Db       2524 CTGGAGGAACTGAGGAAGCAGAGCTCAATGGAAATTCATATGCACGAACTCCGCGGTGCT 2583

Qy       4197 CTGGGTAGCCTCCAAGGAGAAGGCTTTCTTACTGTCCATGGTGACATAGTCAAGAGAGTT 4256
Db       2584 CTTGGCACCCTGATGACTGAAGGCGCGGTGGTTATCCATGGAGACAACGTGAGGAGAGTT 2643

Qy       4257 TGA 4259
Db       2644 TGA 2646
```

RESULT 4
US-10-437-963-44639
; Sequence 44639, Application US/10437963
; Publication No. US20040123343A1
; GENERAL INFORMATION:
; APPLICANT: La Rosa, Thomas J.
; APPLICANT: Kovalic, David K.
; APPLICANT: Zhou, Yihua
; APPLICANT: Cao, Yongwei
; APPLICANT: Wu, Wei
; APPLICANT: Boukharov, Andrey A.
; APPLICANT: Barbazuk, Brad
; APPLICANT: Li, Ping
; TITLE OF INVENTION: Rice Nucleic Acid Molecules and Other Molecules Associated With
; TITLE OF INVENTION: Plants and Uses Thereof for Plant Improvement
; FILE REFERENCE: 38-21(53221)B
; CURRENT APPLICATION NUMBER: US/10/437,963
; CURRENT FILING DATE: 2003-05-14
; NUMBER OF SEQ ID NOS: 204966
; SEQ ID NO 44639
; LENGTH: 3054
; TYPE: DNA
; ORGANISM: Oryza sativa
; FEATURE:
; OTHER INFORMATION: Clone ID: PAT_MRT4530_47680C.1
US-10-437-963-44639

```
  Query Match             19.7%;  Score 857.2;  DB 7;  Length 3054;
  Best Local Similarity   65.3%;  Pred. No. 5.1e-258;
  Matches 1330;  Conservative    0;  Mismatches  658;  Indels   48;  Gaps    3;

Qy       2217 GAGGATACTCTAAATATCGACATGTCAGACATTTATGACCATGATCCTGATCTATACGCA 2276
Db        568 GGGGAGTCGCTCGACGTCAACGCCCACGACGTGTTCGACCACGACCCGGACCTCTACGGC 627

Qy       2277 AAAATTGTTCGATACCCACTCGACATCATCCCCCTGTTGGACACTGAGTGTCAGGAAGTT 2336
Db        628 AAGATGGTCAGGTACCCGCTCGAGGTGCTCGCCATCTTTGACATCGTGCTCATGGACCTC 687

Qy       2337 GCTACCTCTTTACTACCAACGTTTGAGAAGCATATTGAGGCCAGACCTTTCAATCTCAAA 2396
Db        688 GTCGCGCGCATCGAGCCCCTCTTCGAGAAGCACATCCAGACCAGGATCTACAACCTCAAG 747

Qy       2397 GCATCGGTGCACATGCGTGAACTCAACCCTTCAGATATAGACAAATTGGTTTCTGTTAAA 2456
Db        748 TCCTCGGTTTGCTTGAGGAATCTCAACCCTTCTGATATCGAGAAGATGGTGTCCATCAAG 807

Qy       2457 GGAATGGTTATCCGGTGCAGTTCTATCATACCTGAAATTAAGGGGGCCTTCTTCAAATGT 2516
Db        808 GGTATGATAATTCGGTGCAGCTCGGTGATTCCAGAGCTCAAGGAGGCTGTGTTCCGCTGC 867

Qy       2517 TTAGTGTGTGGTCACTCGCCTCCGCTAGTTACAGTTGTTAAAGGGCGGGTTGAGGAGCCA 2576
Db        868 TTGGTTTGCGGGTTCTACTCTGAGCCTGTAATGGTTGATCGAGGGAGGGTTACTGAGCCG 927

Qy       2577 ACAAGGTGTGAAAAGCCAGAATGTGCAGCACGGAATGCTATGTCTCTTATTCACAATCGA 2636
Db        928 CACATCTGTCAGAAAGAACAATGTAAAGCTACAAACTCTATGACTCTTGTGCATAACCGA 987

Qy       2637 TGCACTTTTGCAAATAAGCAGATAGTGCGTCTTCAAGAAACTCCAGATGCCATTCCTGAA 2696
Db        988 TGCAGGTTTGCAGATAAGCAGATCATAAAGTTGCAGGAAACACCAGATGAGATACCAGAA 1047

Qy       2697 GGAGAGACTCCACACACAGTCAGCATGTGTTTATACAACACTATGGTTGATGCTGTGAAG 2756
Db       1048 GGTGGCACTCCACATACTGTTAGTGTCTTGATGCACGATAAGCTTGTAGATGCTGGAAAG 1107

Qy       2757 CCTGGAGATCGTATTGAGGTAACAGGAGTTTTCAAGGCCATGGCAGTTCGAGTTGGTCCG 2816
Db       1108 CCTGGAGACAGGGTTGAGATAACTGGGATATACAGAGCCATGAGTATTAGAGTTGGCCCA 1167

Qy       2817 AATCAACGAACATTACGAGCATTGTATAAGACCTACATCGATTGCGTGCACGTCAAGAAG 2876
Db       1168 ACTCAGAGGACAGTGAAGTCGATTT--------------------------TCAAGAAG 1200

Qy       2877 TCTGACAGGGGTCGACTGCAAACTGAAGATCCTATGGAGATGGATAAGGAGAATGATATG 2936
Db       1201 ACAGACAAGTCTAGACTTCATGTTGAGGACTCCATGGAAACTGATAACCCCAATGCTAAC 1260

Qy       2937 TATGCTGGGTATCATGAAAGTGATACTTCAGAAGCTGCTAATGAAGCAAAGATTCAAAAA 2996
Db       1261 AAGACA------------------ACTGAAGATGATTTTCTCAGAGATAAGGTTGAGAAA 1302

Qy       2997 CTTAAAGAGCTGTCCAAGCTCCCGGGCATTTATGATAGACTTTCAAGGTCGCTGGCTCCA 3056
Db       1303 TTAAAAGAGTTGTCAAAGTTGCCAGATATATATGACAGATTAACTAGGTCATTGGCTCCA 1362

Qy       3057 AGCATTTGGGAGCTTGAAGATATTAAAAAGGGTCTTCTTTGCCAGCTCTTTGGTGGGAAG 3116
Db       1363 AACATATGGGAGCTGGACGACGTTAAAAGGGGCCTCCTTTGCCAGCTTTTTGGTGGAAAT 1422

Qy       3117 GCTAAGAAAATTCCATCTGGAGCATCTTTCCGAGGTGACATCAATGTTTTACTTGTTGGG 3176
Db       1423 GCTTTGAGGCTTCCTTCTGGAGCTAGTTTCCGAGGCGACATCAATATTTGCTTGTTGGT 1482

Qy       3177 GACCCTGGTACCAGTAAATCTCAGCTGCTTCAGTATGTGCACAAGATAGCTCCTCGTGGA 3236
Db       1483 GATCCTGGAACAAGTAAATCCCAGCTTCTCCAATACATGCACAAACTGTCTCCTCGTGGC 1542

Qy       3237 ATCTACACTAGTGGGCGAGGAAGTTCGGCGGTTGGGCTGACAGCGTATGTAACGAAGGAT 3296
Db       1543 ATTTATACAAGTGGCAGAGGAAGTTCAGCTGTTGGCCTTACTGCATACGTTACCAAGGAT 1602

Qy       3297 CCAGAAACTCGAGAGACGGTATTGGAGAGCGGAGCTTTGGTTCTTAGTGATCGTGGGATA 3356
Db       1603 CCTGAAACTGGTGAAACTGTTCTTGAGAGTGGAGCGCTTGTTTTGAGTGACAAAGGTGTT 1662

Qy       3357 TGCTGTATCGATGAGTTCGACAAAATGTCTGATAATGCCCGAAGCATGCTTCATGAGGTA 3416
```

```
Db       1663 TGTTGTATTGATGAATTTGATAAGATGTCTGATAATGCTCGAAGCATGTTACATGAGGTG 1722

Qy       3417 ATGGAGCAACAAACGGTATCTGTAGCCAAAGGGGGTATCATTGCCTCGCTGAACGCTCGG 3476
Db       1723 ATGGAACAACAGACTGTCTCCATTGCCAAGGCTGGAATAATTGCATCTTTAAATGCCAGA 1782

Qy       3477 ACGTCTGTCCTTGCATGTGCAAATCCTAGTGGGTCCCGATACAATGCGCGCCTTTCTGTG 3536
Db       1783 ACATCAGTTCTAGCATGTGCAAATCCTACTGAATCACGTTATAATCCAAGGCTCTCTGTG 1842

Qy       3537 ATTGATAACATCCAGCTTCCTCCAACTCTACTTTCTAGATTTGATTTAATTTACTTAATG 3596
Db       1843 ATTGACAATATCCATCTTCCTCCAACACTGCTTTCTAGGTTTGACCTCATTTATCTGATA 1902

Qy       3597 CTCGACAAACCAGACGAGCAAAACGATCGTCGTCTCGCCAGGCATCTCGTGGCTTTACAC 3656
Db       1903 TTGGACAAGGCAGATGAGCAAACTGATAGACGCCTGGCTAAGCATATTGTTTCGTTGCAT 1962

Qy       3657 TATGAAAACTATGAAGTTTCAAAGCAGGACGCCTTAGATCTACAAACACTTACCGCGTAT 3716
Db       1963 TTTGAGAATCCAAACATAGAGGAGCTCGAGGTCTTGGATTTGCCAACACTTGTAGCCTAC 2022

Qy       3717 ATCACCTATGCTCGTCAGCATGTACATCCTACATTAAGTGATGAAGCTGCTGAAGATTTG 3776
Db       2023 ATAAGTTATGCAAGGAAGCATATACAACCACAGTTATCTGATGAAGCTGCAGAAGAATTG 2082

Qy       3777 ATTAATGGCTATGTTGAGATGCGCCAAAAGGGCAACTTTCCTGGAAGCAGTAAAAAGGTG 3836
Db       2083 ACTCGCGGCTATGTTGAGATGAGGAAAAGAGGAAACAGCCCTGGTAGCAGAAAGAAGGTC 2142

Qy       3837 ATAACAGCCACACCTCGGCAACTCGAAAGTATGATTCGTATCAGTGAAGCCCTAGCTCGA 3896
Db       2143 ATAACTGCGACAGCTCGACAAATTGAGAGCTTGATTCGGCTCAGTGAAGCACTGGCCCGA 2202

Qy       3897 ATGAGATTTTCTGAAGTGGTAGAGAAAGTTGATGCAGCAGAAGCTGTGCGCCTTTTAGAC 3956
Db       2203 ATGCGATTCTCTGAAATGGTTGAAGTACAAGATGTTGTAGAGGCCTTCAGGCTTCTCGAA 2262

Qy       3957 GTCGCTTTGCAGCAATCTGCTACTGATCATGCAACAGGTACGATAGACATGGATCTTATC 4016
Db       2263 GTTGCCATGCAGCAATCGGCAACTGATCATGCCACTGGTACAATCGATATGGATCTTATC 2322

Qy       4017 ACGACTGGAGTGTCGGCCAGCGAGCGTATTCGTCGGGCCAACTTGCTAGCTGCTCTGCGA 4076
Db       2323 ATGACTGGAATATCTGCGAGCGAAAGGCAGAGGCGGGACAATCTCGTTGCAGCAACCCGC 2382

Qy       4077 GAGCTTATAGCAGATAAAATTTCACCTGGCAGCTCCTCTGGCTTGAAGACCAGTCAGCTT 4136
Db       2383 AACCTTGTGATGGAGAAAATGCAGCTTGGAGGGCCCTCAG---TGCGAATGATTGAGTTG 2439

Qy       4137 CTTGAGGATATCCGGAGCCAAAGCAGTGTGGACGTTAGTTTGCAGGATATTAAAAATGCT 4196
Db       2440 CTGGAAGAAATTAGGAAGCAGAGCTCTATGGAAGTTCATCTGCATGATCTTCGCGGTGCT 2499

Qy       4197 CTGGGTAGCCTCCAAGGAGAAGGCTTTCTTACTGTCCATGGTGACATAGTCAAGAG 4252
Db       2500 CTTGGCACTCTGATGACAGAGGGTGCCGTAGTCATCCATGGAGACAGCGTCAAGAG 2555

RESULT 5
US-10-425-115-98898
; Sequence 98898, Application US/10425115
; Publication No. US20040214272A1
; GENERAL INFORMATION:
;  APPLICANT: La Rosa, Thomas J.
;  APPLICANT:  Kovalic, David K.
;  APPLICANT:  Zhou, Yihua
;  APPLICANT:  Cao, Yongwei
;  TITLE OF INVENTION: Nucleic Acid Molecules and Other Molecules Associated With
;  TITLE OF INVENTION:  Plants
;  FILE REFERENCE: 38-21(53222)B
;  CURRENT APPLICATION NUMBER: US/10/425,115
;  CURRENT FILING DATE:  2003-04-28
;  NUMBER OF SEQ ID NOS: 369326
; SEQ ID NO 98898
;   LENGTH: 2186
;   TYPE: DNA
;   ORGANISM: Zea mays
;   FEATURE:
;   NAME/KEY: unsure
;   LOCATION: (1)..(2186)
;   OTHER INFORMATION: unsure at all n locations
;   FEATURE:
;   OTHER INFORMATION: Clone ID: MRT4577_21708C.1
US-10-425-115-98898

  Query Match             15.9%;  Score 690.8;  DB 8;  Length 2186;
  Best Local Similarity   64.7%;  Pred. No. 1.2e-205;
  Matches 1113;  Conservative    0;  Mismatches  537;  Indels   70;  Gaps    3;

Qy       2540 GCTAGTTACAGTTGTTAAAGGGCGGGTTGAGGAGCCAACAAGGTGTGAAAAGCCAGAATG 2599
Db          7 GCCCGTCATGGTTGATAGAGGAAGAGTAACTGAACCGCACATTTGTCAGAAAGAACAATG 66

Qy       2600 TGCAGCACGGAATGCTATGTCTCTTATTCACAATCGATGCACTTTTGCAAATAAGCAGAT 2659
Db         67 TAAAGCCACGAATTCTATGACCCTAGTGCACAACCGATGCAGATTTTCAGACAAGCAGAT 126

Qy       2660 AGTGCGTCTTCAAGAAACTCCAGATGCCATTCCTGAAGGAGAGACTCCACACACAGTCAG 2719
Db        127 CATAAAGTTGCAGGAAACACCAGACGAGATACCAGAAGGTGGCACTCCACATACAGTTAG 186

Qy       2720 CATGTGTTTATACAACACTATGGTTGATGCTGTGAAGCCTGGAGATCGTATTGAGGTAAC 2779
Db        187 TGTCTTGATGCATGATAAGCTTGTTGATGCTGGAAAGCCTGGAGATAGGGTTGAGATAAC 246

Qy       2780 AGGAGTTTTCAAGGCCATGGCAGTTCGAGTTGGTCCGAATCAACGAACATTACGAGCATT 2839
Db        247 TGGAATATACAGAGCTATGAGTATTCGAATTGGACC------------------------ 282

Qy       2840 GTATAAGACCTACATCGATTGCGTGCACGTCAAGAAGTCTGACAGGGGTCGACTGCAAAC 2899
Db        283 ------------------------AACATAAAGAAGACAGACAAGTCTAGGCTTCATGT 317

Qy       2900 TGAAGATCCTATGGAGATGGATAAGGAGAATGATATGTATGCTGGGTATCATGAAAGTGA 2959
Db        318 GGAGGACACCATGGATATTGATAATTCTAACGCTAGCAAATC------------------ 359

Qy       2960 TACTTCAGAAGCTGCTAATGAAGCAAAGATTCAAAAACTTAAAGAGCTGTCCAAGCTCCC 3019
Db        360 TACTGAAGAGGATTTTCTTAGTGATAAGGTTGAGAAACTAAAAGAGCTTTCGAAGTTGCC 419

Qy       3020 GGGCATTTATGATAGACTTTCAAGGTCGCTGGCTCCAAGCATTTGGGAGCTTGAAGATAT 3079
Db        420 TGATATCTATGAAAGATTGACTAGATCATTAGCTCCAAACATATGGGAGTTGGATGATGT 479

Qy       3080 TAAAAAGGGTCTTCTTTGCCAGCTCTTTGGTGGGAAGGCTAAGAAAATTCCATCTGGAGC 3139
Db        480 CAAAAGAGGTCTCCTTTGCCAGCTTTTCGGCGGCAATCCCTTGAAGCTTCCTTCTGGAGC 539

Qy       3140 ATCTTTCCGAGGTGACATCAATGTTTTACTTGTTGGGGACCCTGGTACCAGTAAATCTCA 3199
Db        540 TAGTTTCCGGGGTGACATCAATATTTTACTTGTGGGGGACCCTGGAACAAGTAAATCCCA 599

Qy       3200 GCTGCTTCAGTATGTGCACAAGATAGCTCCTCGTGGAATCTACACTAGTGGGCGAGGAAG 3259
Db        600 GCTTCTCCAGTACATGCATAAACTGTCTCCTCGTGGTATCTATACGAGTGGTAGAGGAAG 659

Qy       3260 TTCGGCGGTTGGGCTGACAGCGTATGTAACGAAGGATCCAGAAACTCGAGAGACGGTATT 3319
Db        660 TTCTGCTGTTGGTCTTACTGCTTATGTTACCAAAGACCCTGAGACTGGCGAAACTGTTCT 719

Qy       3320 GGAGAGCGGAGCTTTGGTTCTTAGTGATCGTGGGATATGCTGTATCGATGAGTTCGACAA 3379
Db        720 AGAAAGTGGAGCACTTGTTTTGAGTGACAAAGGTGTTTGTTGCATAGATGAGTTTGATAA 779
```

```
Qy       3380 AATGTCTGATAATGCCCGAAGCATGCTTCATGAGGTAATGGAGCAACAAACGGTATCTGT 3439
Db        780 GATGTCTGATAATGCCCGAAGCATGTTACACGAGGTGATGGAACAGCAGACAGTATCCAT 839

Qy       3440 AGCCAAAGGGGGTATCATTGCCTCGCTGAACGCTCGGACGTCTGTCCTTGCATGTGCAAA 3499
Db        840 TGCGAAGGCTGGAATAATTGCATCTTTAAACGCTAGGACATCTGTCCTGGCATGTGCCAA 899

Qy       3500 TCCTAGTGGGTCCCGATACAATGCGCGCCTTTCTGTGATTGATAACATCCAGCTTCCTCC 3559
Db        900 TCCTACTGAATCACGTTACAATCCAAGGCTCTCTGTAATTGACAACATCCACTTAGCGCC 959

Qy       3560 AACTCTACTTTCTAGATTTGATTTAATTTACTTAATGCTCGACAAACCAGACGAGCAAAA 3619
Db        960 AACGCTACTTTCAAGATTCGACCTGATTTATCTTATCTTGGACAAGGCGGATGAGCAAAC 1019

Qy       3620 CGATCGTCGTCTCGCCAGGCATCTCGTGGCTTTACACTATGAAAACTATGAAGTTTCAAA 3679
Db       1020 TGATAGGCGCCTGGCAAAGCATATTGTTTCGTTGCATTTTGAGAATCCAAATTTAGAGGA 1079

Qy       3680 GCAGGACGCCTTAGATCTACAAACACTTACCGCGTATATCACCTATGCTCGTCAGCATGT 3739
Db       1080 GCTCGAGGTCTTGGACTTGCAGACACTAGTTTCCTACATAAGCTATGCAAGGAAGTATAT 1139

Qy       3740 ACATCCTACATTAAGTGATGAAGCTGCTGAAGATTTGATTAATGGCTATGTTGAGATGCG 3799
Db       1140 TCAGCCACAGTTATCTGATGAAGCTGCAGAAGAGTTAACTCGTGGCTATGTCGAGATGAG 1199

Qy       3800 CCAAAAGGGCAACTTTCCTGGAAGCAGTAAAAAGGTGATAACAGCCACACCTCGGCAACT 3859
Db       1200 AAAAAGAGGGAATAGCCCTGGGAGCAGAAAGAAGGTCATAACAGCAACCGCTAGACAAAT 1259

Qy       3860 CGAAAGTATGATTCGTATCAGTGAAGCCCTAGCTCGAATGAGATTTTCTGAAGTGGTAGA 3919
Db       1260 AGAGAGTTTGATCCGTCTCAGCGAAGCATTAGCCCGAATGCGGTTCTCTGAAGTGGTCGA 1319

Qy       3920 GAAAGTTGATGCAGCAGAAGCTGTGCGCCTTTTAGACGTCGCTTTGCAGCAATCTGCTAC 3979
Db       1320 GGTGCGGGATGTTGTGGAGGCATTCAGGCTTCTTGAAGTCGCCATGCAGCAGTCTGCGAC 1379

Qy       3980 TGATCATGCAACAGGTACGATAGACATGGATCTTATCACGACTGGAGTGTCGGCCAGCGA 4039
Db       1380 GGATCATGCAACTGGGACGATTGATATGGATCTGATCATGACGGGGATATCCGCAAGCGA 1439

Qy       4040 GCGTATTCGTCGGGCCAACTTGCTAGCTGCTCTGCGAGAGCTTATAGCAGATAAAATTTC 4099
Db       1440 AAGGCAGAGGCGGGAGAACCTCGTTGCCGCAACCCGTAACCTGATTGCGGAGAAAATGCA 1499

Qy       4100 ACCTGGCAGCTCCTCTGGCTTGAAGACCAGTCAGCTTCTTGAGGATATCCGGAGCCAAAG 4159
Db       1500 GCTTGGAGGCCCCTC---GATGCGCATGATTGAGTTGCTGGAGGAACTGAGGAAGCAGAG 1556

Qy       4160 CAGTGTGGACGTTAGTTTGCAGGATATTAAAAATGCTCTGGGTAGCCTCCAAGGAGAAGG 4219
Db       1557 CTCAATGGAAATTCATATGCACGAACTCCGCGGTGCTCTTGGCACCCTGATGACTGAAGG 1616

Qy       4220 CTTTCTTACTGTCCATGGTGACATAGTCAAGAGAGTTTGA 4259
Db       1617 CGCGGTGGTTATCCATGGAGACAACGTGAGGAGAGTTTGA 1656
```

RESULT 6
US-09-828-062-2
; Sequence 2, Application US/09828062
; Publication No. US20030097675A1
; GENERAL INFORMATION:
; APPLICANT: COSTA E SILVA, OSWALDO DA
; APPLICANT: BOHNERT, HANS J.
; APPLICANT: VAN THIELEN, NOCHA
; APPLICANT: CHEN, ROUYING
; APPLICANT: SARRIA-MILLAN, RODRIGO
; TITLE OF INVENTION: CELL CYCLE STRESS-RELATED PROTEINS AND METHODS OF USE
; TITLE OF INVENTION: IN PLANTS
; FILE REFERENCE: 16313-0031
; CURRENT APPLICATION NUMBER: US/09/828,062
; CURRENT FILING DATE: 2001-08-20
; PRIOR APPLICATION NUMBER: 60/196,001
; PRIOR FILING DATE: 2000-04-07
; NUMBER OF SEQ ID NOS: 34
; SOFTWARE: PatentIn Ver. 2.1
; SEQ ID NO 2
; LENGTH: 723
; TYPE: DNA
; ORGANISM: Physcomitrella patens
US-09-828-062-2

```
  Query Match             15.5%;  Score 673;  DB 3;  Length 723;
  Best Local Similarity   97.1%;  Pred. No. 2.3e-200;
  Matches  696;  Conservative    0;  Mismatches   20;  Indels    1;  Gaps    1;

Qy       2753 GAAGCCTGGAGATCGTATTGAGGTAACAGGAGTTTTCAAGGCCATGGCAGTTCGAGTTGG 2812
Db          3 GCACCAGGGAGATCGTATTGAGGTAACAGGAGTTTTCAAGGCCATGGCAGTTCGAGTTGG 62

Qy       2813 TCCGAATCAACGAACATTACGAGCATTGTATAAGACCTACATCGATTGCGTGCACGTCAA 2872
Db         63 TCCGAATCAACGAACATTACGAGCATTGTATAAGACCTACATCGATTGCGTGCACGTCAA 122

Qy       2873 GAAGTCTGACAGGGGTCGACTGCAAACTGAAGATCCTATGGAGATGGATAAGGAGAATGA 2932
Db        123 GAAGTCTGACAGGGGTCGACTGCAAACTGAAGATCCTATGGAGATGGATAAGGAGAATGA 182

Qy       2933 TATGTATGCTGGGTATCATGAAAGTGATACTTCAGAAGCTGCTAATGAAGCAAAGATTCA 2992
Db        183 TATGTATGCTGGGTATCATGAAAGTGATACTTCAGAAGCTGCTAATGAAGCAAAGATTCA 242

Qy       2993 AAAACTTAAAGAGCTGTCCAAGCTCCCGGGCATTTATGATAGACTTTCAAGGTCGCTGGC 3052
Db        243 AAAACTTAAAGAGCTGTCCAAGCTCCCGGACATTTATGATAGACTTTCAAGGTCGCTGGC 302

Qy       3053 TCCAAGCATTTGGGAGCTTGAAGATATTAAAAAGGGTCTTCTTTGCCAGCTCTTTGGTGG 3112
Db        303 TCCAAGCATTTGGGAGCTTGAAGATATTAAAAAGGGTCTTCTTTGCCAGCTCTTTGGTGG 362

Qy       3113 GAAGGCTAAGAAAATTCCATCTGGAGCATCTTTCCGAGGTGACATCAATGTTTTACTTGT 3172
Db        363 GAAGGCTAAGAAAATTCCATCTGGAGCATCTTTCCGAGGTGACATCAATGTTTTACTTGT 422

Qy       3173 TGGGGACCCTGGTACCAGTAAATCTCAGCTGCTTCAGTATGTGCACAAGATAGCTCCTCG 3232
Db        423 TGGGGACCCTGGTACCAGTAAATCTCAGCTGCTTCAGTATGTGCACAAGATAGCTCCTCG 482

Qy       3233 TGGAATCTACACTAGTGGGCGAGGAAGTTCGGCGGTTGGGCTGACAGCGTATGT-AACGA 3291
Db        483 TGGAATCTACACTAGTGGGCGAGGAAGTTCGGCGGTTGGGCTGACAGCGTATGTAAACGA 542

Qy       3292 AGGATCCAGAAACTCGAGAGACGGTATTGGAGAGCGGAGCTTTGGTTCTTAGTGATCGTG 3351
Db        543 AGGATCCAGAAACTCGAGAGACGGTATTGGAGAGCGGAGCTTTGGTTCTTAGTGATCGTG 602

Qy       3352 GGATATGCTGTATCGATGAGTTCGACAAAATGTCTGATAATGCCCGAAGCATGCTTCATG 3411
Db        603 GGATATGCTGTATCGATGAGTTCGACAAAATGTCTGATAATGCCCGAAGCATGCTTCATG 662

Qy       3412 AGGTAATGGAGCAACAAACGGTATCTGTAGCCAAAGGGGGTATCATTGCCTCGCTGA 3468
Db        663 AGGTAATGGAGCAACAAACGGTATCTGGACCCAAGCGGTTCATGCTCGTGAAGCCGA 719
```

RESULT 7
US-10-768-511-2
; Sequence 2, Application US/10768511
; Publication No. US20040128721A1
; GENERAL INFORMATION:
; APPLICANT: COSTA E SILVA, OSWALDO DA

```
;  APPLICANT:  BOHNERT, HANS J.
;  APPLICANT:  VAN THIELEN, NOCHA
;  APPLICANT:  CHEN, ROUYING
;  APPLICANT:  SARRIA-MILLAN, RODRIGO
;  TITLE OF INVENTION: CELL CYCLE STRESS-RELATED PROTEINS AND METHODS OF USE
;  TITLE OF INVENTION:  IN PLANTS
;  FILE REFERENCE: 16313-0031
;  CURRENT APPLICATION NUMBER: US/10/768,511
;  CURRENT FILING DATE:  2004-01-30
;  PRIOR APPLICATION NUMBER: 60/196,001
;  PRIOR FILING DATE: 2000-04-07
;  NUMBER OF SEQ ID NOS: 34
;  SOFTWARE: PatentIn Ver. 2.1
; SEQ ID NO 2
;   LENGTH: 723
;   TYPE: DNA
;   ORGANISM: Physcomitrella patens
US-10-768-511-2

  Query Match             15.5%;  Score 673;  DB 7;  Length 723;
  Best Local Similarity   97.1%;  Pred. No. 2.3e-200;
  Matches  696;  Conservative    0;  Mismatches   20;  Indels    1;  Gaps    1;

Qy       2753 GAAGCCTGGAGATCGTATTGAGGTAACAGGAGTTTTCAAGGCCATGGCAGTTCGAGTTGG 2812
Db          3 GCACCAGGGAGATCGTATTGAGGTAACAGGAGTTTTCAAGGCCATGGCAGTTCGAGTTGG 62

Qy       2813 TCCGAATCAACGAACATTACGAGCATTGTATAAGACCTACATCGATTGCGTGCACGTCAA 2872
Db         63 TCCGAATCAACGAACATTACGAGCATTGTATAAGACCTACATCGATTGCGTGCACGTCAA 122

Qy       2873 GAAGTCTGACAGGGGTCGACTGCAAACTGAAGATCCTATGGAGATGGATAAGGAGAATGA 2932
Db        123 GAAGTCTGACAGGGGTCGACTGCAAACTGAAGATCCTATGGAGATGGATAAGGAGAATGA 182

Qy       2933 TATGTATGCTGGGTATCATGAAAGTGATACTTCAGAAGCTGCTAATGAAGCAAAGATTCA 2992
Db        183 TATGTATGCTGGGTATCATGAAAGTGATACTTCAGAAGCTGCTAATGAAGCAAAGATTCA 242

Qy       2993 AAAACTTAAAGAGCTGTCCAAGCTCCCGGGCATTTATGATAGACTTTCAAGGTCGCTGGC 3052
Db        243 AAAACTTAAAGAGCTGTCCAAGCTCCCGGACATTTATGATAGACTTTCAAGGTCGCTGGC 302

Qy       3053 TCCAAGCATTTGGGAGCTTGAAGATATTAAAAAGGGTCTTCTTTGCCAGCTCTTTGGTGG 3112
Db        303 TCCAAGCATTTGGGAGCTTGAAGATATTAAAAAGGGTCTTCTTTGCCAGCTCTTTGGTGG 362

Qy       3113 GAAGGCTAAGAAAATTCCATCTGGAGCATCTTTCCGAGGTGACATCAATGTTTTACTTGT 3172
Db        363 GAAGGCTAAGAAAATTCCATCTGGAGCATCTTTCCGAGGTGACATCAATGTTTTACTTGT 422

Qy       3173 TGGGGACCCTGGTACCAGTAAATCTCAGCTGCTTCAGTATGTGCACAAGATAGCTCCTCG 3232
Db        423 TGGGGACCCTGGTACCAGTAAATCTCAGCTGCTTCAGTATGTGCACAAGATAGCTCCTCG 482

Qy       3233 TGGAATCTACACTAGTGGGCGAGGAAGTTCGGCGGTTGGGCTGACAGCGTATGT-AACGA 3291
Db        483 TGGAATCTACACTAGTGGGCGAGGAAGTTCGGCGGTTGGGCTGACAGCGTATGTAAACGA 542

Qy       3292 AGGATCCAGAAACTCGAGAGACGGTATTGGAGAGCGGAGCTTTGGTTCTTAGTGATCGTG 3351
Db        543 AGGATCCAGAAACTCGAGAGACGGTATTGGAGAGCGGAGCTTTGGTTCTTAGTGATCGTG 602

Qy       3352 GGATATGCTGTATCGATGAGTTCGACAAAATGTCTGATAATGCCCGAAGCATGCTTCATG 3411
Db        603 GGATATGCTGTATCGATGAGTTCGACAAAATGTCTGATAATGCCCGAAGCATGCTTCATG 662

Qy       3412 AGGTAATGGAGCAACAAACGGTATCTGTAGCCAAAGGGGGTATCATTGCCTCGCTGA 3468
Db        663 AGGTAATGGAGCAACAAACGGTATCTGGACCCAAGCGGTTCATGCTCGTGAAGCCGA 719

RESULT 8
US-10-425-114-19722
; Sequence 19722, Application US/10425114
; Publication No. US20040034888A1
; GENERAL INFORMATION:
;  APPLICANT: Liu, Jingdong
;  APPLICANT:  Zhou, Yihua
;  APPLICANT:  Kovalic, David K.
;  APPLICANT:  Screen, Steven E
;  APPLICANT:  Tabaska, Jack E
;  APPLICANT:  Cao, Yongwei
;  TITLE OF INVENTION: Nucleic Acid Molecules and Other Molecules Associated With
;  TITLE OF INVENTION:  Plants and Uses Thereof for Plant Improvement
;  FILE REFERENCE: 38-21(53313)B
;  CURRENT APPLICATION NUMBER: US/10/425,114
;  CURRENT FILING DATE:  2003-04-28
;  NUMBER OF SEQ ID NOS: 73128
; SEQ ID NO 19722
;   LENGTH: 1571
;   TYPE: DNA
;   ORGANISM: Zea mays
;   FEATURE:
;   OTHER INFORMATION: Clone ID: LIB3150-112-C5_FLI
US-10-425-114-19722

  Query Match             14.6%;  Score 634.2;  DB 7;  Length 1571;
  Best Local Similarity   67.5%;  Pred. No. 6.6e-188;
  Matches  907;  Conservative    0;  Mismatches  433;  Indels    3;  Gaps    1;

Qy       2917 TGGATAAGGAGAATGATATGTATGCTGGGTATCATGAAAGTGATACTTCAGAAGCTGCTA 2976
Db         32 TGGAGGACACCATGGATATTGATAATTCTAACGCTAGCAAATCTACTGAAGAGGATTTTC 91

Qy       2977 ATGAAGCAAAGATTCAAAAACTTAAAGAGCTGTCCAAGCTCCCGGGCATTTATGATAGAC 3036
Db         92 TTAGTGATAAGGTTGAGAAACTAAAAGAGCTTTCGAAGTTGCCTGATATCTATGAAAGAT 151

Qy       3037 TTTCAAGGTCGCTGGCTCCAAGCATTTGGGAGCTTGAAGATATTAAAAAGGGTCTTCTTT 3096
Db        152 TGACTAGATCATTAGCTCCAAACATATGGGAGTTGGATGATGTCAAAAGAGGTCTCCTTT 211

Qy       3097 GCCAGCTCTTTGGTGGAAGGCTAAGAAAATTCCATCTGGAGCATCTTTCCGAGGTGACA 3156
Db        212 GCCAGCTTTTCGGCGGCAATCCCTTGAAGCTTCCTTCTGGAGCTAGTTTCCGGGGTGACA 271

Qy       3157 TCAATGTTTTACTTGTTGGGGACCCTGGTACCAGTAAATCTCAGCTGCTTCAGTATGTGC 3216
Db        272 TCAATATTTTACTTGTGGGGGACCCTGGAACAAGTAAATCCCAGCTTCTCCAGTACATGC 331

Qy       3217 ACAAGATAGCTCCTCGTGGAATCTACACTAGTGGGCGAGGAAGTTCGGCGGTTGGGCTGA 3276
Db        332 ATAAACTGTCTCCTCGTGGTATCTATACGAGTGGTAGAGGAAGTTCTGCTGTTGGTCTTA 391

Qy       3277 CAGCGTATGTAACGAAGGATCCAGAAACTCGAGAGACGGTATTGGAGAGCGGAGCTTTGG 3336
Db        392 CTGCTTATGTTACCAAAGACCCTGAGACTGGCGAAACTGTTCTAGAAAGTGGAGCACTTG 451

Qy       3337 TTCTTAGTGATCGTGGGATATGCTGTATCGATGAGTTCGACAAAATGTCTGATAATGCCC 3396
Db        452 TTTTGAGTGACAAAGGTGTTTGTTGCGTAGATGAGTTTGATAAGATGTCTGATAATGCCC 511

Qy       3397 GAAGCATGCTTCATGAGGTAATGGAGCAACAAACGGTATCTGTAGCCAAAGGGGGTATCA 3456
Db        512 GAAGCATGTTACACGAGGTGATGGAACAGCAGACAGTATCCATTGCGAAGGCTGGAATAA 571

Qy       3457 TTGCCTCGCTGAACGCTCGGACGTCTGTCCTTGCATGTGCAAATCCTAGTGGGTCCCGAT 3516
Db        572 TTGCATCTTTAAACGCTAGGACATCTGTCCTGGCATGTGCCAATCCTACTGAATCACGTT 631

Qy       3517 ACAATGCGCGCCTTTCTGTGATTGATAACATCCAGCTTCCTCCAACTCTACTTTCTAGAT 3576
Db        632 ACAATCCAAGGCTCTCTGTAATTGACAACATCCACTTAGCGCCAACGCTACTTTCAAGAT 691

Qy       3577 TTGATTTAATTTACTTAATGCTCGACAAACCAGACGAGCAAAACGATCGTCGTCTCGCCA 3636
```

```
Db      692 TCGACCTGATTTATCTTATCTTGGACAAGGCGGATGAGCAAACTGATAGGCGCCTGGCAA 751

Qy     3637 GGCATCTCGTGGCTTTACACTATGAAAACTATGAAGTTTCAAAGCAGGACGCCTTAGATC 3696
Db      752 AGCATATTGTTTCGTTGCATTTTGAGAATCCAAATTTAGAGGAGCTCGAGGTCTTGGACT 811

Qy     3697 TACAAACACTTACCGCGTATATCACCTATGCTCGTCAGCATGTACATCCTACATTAAGTG 3756
Db      812 TGCAGACACTAGTTTCCTACATAAGCTATGCAAGGAAGTATATTCAGCCACAGTTATCTG 871

Qy     3757 ATGAAGCTGCTGAAGATTTGATTAATGGCTATGTTGAGATGCGCCAAAAGGGCAACTTTC 3816
Db      872 ATGAAGCTGCAGAAGAGTTAACTCGTGGCTATGTCGAGATGAGAAAAAGAGGGAATAGCC 931

Qy     3817 CTGGAAGCAGTAAAAAGGTGATAACAGCCACACCTCGGCAACTCGAAAGTATGATTCGTA 3876
Db      932 CTGGGAGCAGAAAGAAGGTCATAACAGCAACCGCTAGACAAATAGAGAGTTTGATCCGTC 991

Qy     3877 TCAGTGAAGCCCTAGCTCGAATGAGATTTTCTGAAGTGGTAGAGAAAGTTGATGCAGCAG 3936
Db      992 TCAGCGAAGCATTAGCCCGAATGCGGTTCTCTGAAGTGGTCGAGGTGCGGGATGTTGTGG 1051

Qy     3937 AAGCTGTGCGCCTTTTAGACGTCGCTTTGCAGCAATCTGCTACTGATCATGCAACAGGTA 3996
Db     1052 AGGCATTCAGGCTTCTTGAAGTCGCCATGCAGCAGTCTGCGACGGATCATGCAACTGGTA 1111

Qy     3997 CGATAGACATGGATCTTATCACGACTGGAGTGTCGGCCAGCGAGCGTATTCGTCGGGCCA 4056
Db     1112 CGATTGATATGGATCTGATCATGACGGGGATATCCGCAAGCGAAAGGCAGAGGCGGGAGA 1171

Qy     4057 ACTTGCTAGCTGCTCTGCGAGAGCTTATAGCAGATAAAATTTCACCTGGCAGCTCCTCTG 4116
Db     1172 ACCTCGTTGCCGCAACCCGTAACCTGATTGCGGAGAAAATGCAGCTTGGAGGCCCCTC-- 1229

Qy     4117 GCTTGAAGACCAGTCAGCTTCTTGAGGATATCCGGAGCCAAAGCAGTGTGGACGTTAGTT 4176
Db     1230 -GATGCGCATGATTGAGTTGCTGGAGGAACTGAGGAAGCAGAGCTCAATGGAAATTCATA 1288

Qy     4177 TGCAGGATATTAAAAATGCTCTGGGTAGCCTCCAAGGAGAAGGCTTTCTTACTGTCCATG 4236
Db     1289 TGCACGAACTCCGCGGTGCTCTTGGCACCCTGATGACTGAAGGCGCGGTGGTTATCCATG 1348

Qy     4237 GTGACATAGTCAAGAGAGTTTGA 4259
Db     1349 GAGACAACGTGAGGAGAGTTTGA 1371
```

RESULT 9
US-10-425-115-98897
; Sequence 98897, Application US/10425115
; Publication No. US20040214272A1
; GENERAL INFORMATION:
; APPLICANT: La Rosa, Thomas J.
; APPLICANT: Kovalic, David K.
; APPLICANT: Zhou, Yihua
; APPLICANT: Cao, Yongwei
; TITLE OF INVENTION: Nucleic Acid Molecules and Other Molecules Associated With
; TITLE OF INVENTION: Plants
; FILE REFERENCE: 38-21(53222)B
; CURRENT APPLICATION NUMBER: US/10/425,115
; CURRENT FILING DATE: 2003-04-28
; NUMBER OF SEQ ID NOS: 369326
; SEQ ID NO 98897
; LENGTH: 2939
; TYPE: DNA
; ORGANISM: Zea mays
; FEATURE:
; OTHER INFORMATION: Clone ID: MRT4577_21707C.1
US-10-425-115-98897

```
  Query Match             13.8%;  Score 598.6;  DB 8;  Length 2939;
  Best Local Similarity   56.7%;  Pred. No. 1.7e-176;
  Matches 1158;  Conservative    0;  Mismatches  864;  Indels   21;  Gaps    2;

Qy     2217 GAGGATACTCTAAATATCGACATGTCAGACATTTATGACCATGATCCTGATCTATACGCA 2276
Db      638 GGGGAGTCGCTCGACGTCGATGCGCACGACGTGTTTGACCACGACCCAGACCTCTACAGC 697

Qy     2277 AAAATTGTTCGATACCCACTCGACATCATCCCCCTGTTGGACACTGAGTGTCAGGAAGTT 2336
Db      698 AAGATGGTTCGCTATCCGCTCGAGGTGCTCGCCATCTTCGACATCGTGCTCATGGACCTC 757

Qy     2337 GCTACCTCTTTACTACCAACGTTTGAGAAGCATATTGAGGCCAGACCTTTCAATCTCAAA 2396
Db      758 GTCGCGCGCATCGAGCCGCTCTTCGAGAAGCACATCCAGACCAGGATCTACAACCTCAAG 817

Qy     2397 GCATCGGTGCACATGCGTGAACTCAACCCTTCAGATATAGACAAATTGGTTTCTGTTAAA 2456
Db      818 TCGTCCATTTGCTTGAGGAATCTCAACCCATCTGATATTGAGAAGATGGTATCCATCAAG 877

Qy     2457 GGAATGGTTATCCGGTGCAGTTCTATCATACCTGAAATTAAGGGGGCCTTCTTCAAATGT 2516
Db      878 GGTATGATAATTAGATGCAGCTCGGTCATACCGGAGCTCAAGGAGGCTGTGTTCCGCTGC 937

Qy     2517 TTAGTGTGTGGTCACTCGCCTCCGCTAGTTACAGTTGTTAAAGGGCGGGTTGAGGAGCCA 2576
Db      938 CTGGTTTGTGGTTTCTACTCAGAGCCCGTCATGGTTGATAGAGGAAGAGTAACTGAACCA 997

Qy     2577 ACAAGGTGTGAAAAGCCAGAATGTGCAGCACGGAATGCTATGTCTCTTATTCACAATCGA 2636
Db      998 CACATTTGTCAGAAAGAACAATGTAAAGCCACAAATTCTATGACCCTAGTGCACAACAGA 1057

Qy     2637 TGCACTTTTGCAAATAAGCAGATAGTGCGTCTTCAAGAAACTCCAGATGCCATTCCTGAA 2696
Db     1058 TGCAGATTTTCAGACAAGCAGATCATAAAGTTGCAGGAAACACCAGACGAGATACCAGAA 1117

Qy     2697 GGAGAGACTCCACACACAGTCAGCATGTGTTTATACAACACTATGGTTGATGCTGTGAAG 2756
Db     1118 GGTGGCACTCCACATACAGTTAGTGTCTTGATGCATGATAAGCTTGTTGATGCTGGAAAG 1177

Qy     2757 CCTGGAGATCGTATTGAGGTAACAGGAGTTTTCAAGGCCATGGCAGTTCGAGTTGGTCCG 2816
Db     1178 CCTGGAGATAGGGTTGAGATAACTGGAATATACAGAGCTATGAGTATTCGAATTGGACCA 1237

Qy     2817 AATCAACGAACATTACGAGCATTGTATAAGACCTACATCGATTGCGTGCACGTCAAGAAG 2876
Db     1238 ACTCAAAGGACAGTGAAGTCTATATTCAAGACATATATTGCCCGGATGCGGTTCTCTGAA 1297

Qy     2877 TCTGACAGGGGTCGACTGCAAACTGAAGATCCTATGGAGATGGATAAGGAGAATGATATG 2936
Db     1298 GTGGTCGAGGTGCGAGATGTTGTGGAGGCCTTCAGGCTTCTTGAAGTTGCCATGCAGCAG 1357

Qy     2937 TATGCTGGGTATCATGAAAGTGATACTTCAGAAGCTGCTAATGAAGCAAAGATTCAAAAA 2996
Db     1358 TCTGCAACGGATCATGCAACTGGGACGATTGATATGGATCTAATCATGACGGGGATATCT 1417

Qy     2997 CTTAAAGAGCTGTCCAAGCTCCCGGGCATTTATGATAGACTTTCAAGGTCGCTGGCTCCA 3056
Db     1418 GCAAGTGAAAGGCAGAGACGGAACGACCTTGTTGCT------------------GCAACC 1459

Qy     3057 AGCATTTGGGAGCTTGAAGATATTAAAAAGGGTCTTCTTTGCCAGCTCTTTGGTGGGAAG 3116
Db     1460 CGTAACCTGATCGTGGAGAAAATGCAGCTCGGAGGGCCTTCAATGCATATGATTGAGTTG 1519

Qy     3117 GCTAAGAAAATTCCATCTGGAGCATCTTTCCGAGGTGACATCAATGTTTTACTTGTTGGG 3176
Db     1520 CTGTAAGAGCTCAGGTCGCGGAGCTGAATGGAGAGCCATCTGCACGAATTCCTCGGGGCG 1579

Qy     3177 GACCCTGGTACCAGTAAATCTCAGCTGCTTCAGTATGTGCACAAGATAGCTCCTCGTGGA 3236
Db     1580 CTTGGCACTATGATGACAGAAGGTGTGCTTCTGATCCATGGAAAAAAGGTGAAGAGAGGT 1639

Qy     3237 ATCTACACTAGTGGGCGAGGAAGTTCGGCGGTTGGGCTGACAGCGTATGTAACGAAGGAT 3296
Db     1640 TGATGTTAACTGCCATCTGCAGGGATGGTTGTGTGAGCTCCGTCTACCGTTTCCTTCAAC 1699
```

```
Qy   3297 CCAGAAACTCGAGAGACGGTATTGGAGAGCGGAGCTTTGGTTCTTAGTGATCGTGGGATA 3356
Db   1700 AGTCAAGGTTGCGAAGATCTGATCGACTGTAGGGCATTTGCTTTCAGAGTGCTAAGTAGT 1759

Qy   3357 TGCTGTATCGATGAGTTCGACAAAATGTCTGATAATGCCCGAAGCATGCTTCATGAGGTA 3416
Db   1760 TTGTGCTGAGATGAGTTTCATAAGATGTCAGTTAATGCCCGAAGCATGTTACACGAGGTG 1819

Qy   3417 ATGGAGCAACAAACGGTATCTGTAGCCAAAGGGGGTATCATTGCCTCGCTGAACGCTCGG 3476
Db   1820 ATGGAACAGCAGACAGTATCCATTGCGAAGGCTGGAATAATTGCATCTTTAAACGCTAGG 1879

Qy   3477 ACGTCTGTCCTTGCATGTGCAAATCCTAGTGGGTCCCGATACAATGCGCGCCTTTCTGTG 3536
Db   1880 ACATCTGTCCTGGCATGTGCCAATCCTACTGAATCACGTTACAATCCAAGGCTCTCTGTA 1939

Qy   3537 ATTGATAACATCCAGCTTCCTCCAACTCTACTTTCTAGATTTGATTTAATTTACTTAATG 3596
Db   1940 ATTGACAACATCCACTTAGCGCCAACGCTACTTTCAAGATTCGACCTGATTTATCTTATC 1999

Qy   3597 CTCGACAAACCAGACGAGCAAAACGATCGTCGTCTCGCCAGGCATCTCGTGGCTTTACAC 3656
Db   2000 TTGGACAAGGCGGATGAGCAAACTGATAGGCGCCTGGCAAAGCATATTGTTTCGTTGCAT 2059

Qy   3657 TATGAAAACTATGAAGTTTCAAAGCAGGACGCCTTAGATCTACAAACACTTACCGCGTAT 3716
Db   2060 TTTGAGAATCCAAATTTAGAGGAGCTCGAGGTCTTGGACTTGCAGACACTAGTTTCCTAC 2119

Qy   3717 ATCACCTATGCTCGTCAGCATGTACATCCTACATTAAGTGATGAAGCTGCTGAAGATTTG 3776
Db   2120 ATAAGCTATGCAAGGAAGTATATTCAGCCACAGTTATCTGATGAAGCTGCAGAAGAGTTA 2179

.Qy  3777 ATTAATGGCTATGTTGAGATGCGCCAAAAGGGCAACTTTCCTGGAAGCAGTAAAAAGGTG 3836
Db   2180 ACTCGTGGCTATGTCGAGATGAGAAAAAGAGGGAATAGCCCTGGGAGCAGAAAGAAGGTC 2239

Qy   3837 ATAACAGCCACACCTCGGCAACTCGAAAGTATGATTCGTATCAGTGAAGCCCTAGCTCGA 3896
Db   2240 ATAACAGCAACCGCTAGACAAATAGAGAGTTTGATCCGTCTCAGCGAAGCATTAGCCCGA 2299

Qy   3897 ATGAGATTTTCTGAAGTGGTAGAGAAAGTTGATGCAGCAGAAGCTGTGCGCCTTTTAGAC 3956
Db   2300 ATGCGGTTCTCTGAAGTGGTCGAGGTGCGGGATGTTGTGGAGGCATTCAGGCTTCTTGAA 2359

Qy   3957 GTCGCTTTGCAGCAATCTGCTACTGATCATGCAACAGGTACGATAGACATGGATCTTATC 4016
Db   2360 GTCGCCATGCAGCAGTCTGCGACGGATCATGCAACTGGTACGATTGATATGGATCTGATC 2419

Qy   4017 ACGACTGGAGTGTCGGCCAGCGAGCGTATTCGTCGGGCCAACTTGCTAGCTGCTCTGCGA 4076
Db   2420 ATGACGGGGATATCCGCAAGCGAAAGGCAGAGGCGGGAGAACCTCGTTGCCGCAACCCGT 2479

Qy   4077 GAGCTTATAGCAGATAAAATTTCACCTGGCAGCTCCTCTGGCTTGAAGACCAGTCAGCTT 4136
Db   2480 AACCTGATTGCGGAGAAAATGCAGCTTGGAGGCCCCTC---GATGCGCATGATTGAGTTG 2536

Qy   4137 CTTGAGGATATCCGGAGCCAAAGCAGTGTGGACGTTAGTTTGCAGGATATTAAAAATGCT 4196
Db   2537 CTGGAGGAACTGAGGAAGCAGAGCTCAATGGAAATTCATATGCACGAACTCCGCGGTGCT 2596

Qy   4197 CTGGGTAGCCTCCAAGGAGAAGGCTTTCTTACTGTCCATGGTGACATAGTCAAGAGAGTT 4256
Db   2597 CTTGGCACCCTGATGACTGAAGGCGCGGTGGTTATCCATGGAGACAACGTGAGGAGAGTT 2656

Qy   4257 TGA 4259
Db   2657 TGA 2659
```

RESULT 10
US-10-320-797-2332
; Sequence 2332, Application US/10320797
; Publication No. US20040014955A1
; GENERAL INFORMATION:
; APPLICANT: Eroshkin, Alexey M.
; APPLICANT: Zamudio, Carlos
; TITLE OF INVENTION: IDENTIFICATION OF ESSENTIAL GENES OF CRYPTOCOCCUS NEOFORMANS AND
; TITLE OF INVENTION: METHODS OF USE
; FILE REFERENCE: 10182-021-999
; CURRENT APPLICATION NUMBER: US/10/320,797
; CURRENT FILING DATE: 2002-12-16
; PRIOR APPLICATION NUMBER: 60/341,261
; PRIOR FILING DATE: 2001-12-17
; NUMBER OF SEQ ID NOS: 3361
; SOFTWARE: PatentIn version 3.1
; SEQ ID NO 2332
;   LENGTH: 2970
;   TYPE: DNA
;   ORGANISM: Cryptococcus neoformans
US-10-320-797-2332

```
  Query Match             11.3%;  Score 492.8;  DB 7;  Length 2970;
  Best Local Similarity   57.1%;  Pred. No. 3.8e-143;
  Matches  993;  Conservative    0;  Mismatches  717;  Indels   30;  Gaps    4;

Qy   2372 TGAGGCCAGACCTTTCAATCTCAAAGCATCGGTGCACATGCGTGAACTCAACCCTTCAGA 2431
Db    768 TAAAGTCAGACCGTTTGGTGGAGAGAAGACGGTCAACATGAGGGATCTTAATCCCGGAGA 827

Qy   2432 TATAGACAAATTGGTTTCTGTTAAAGGAATGGTTATCCGGTGCAGTTCTATCATACCTGA 2491
Db    828 TACTGACAAGCTGGTGACCGTGAAGGGTCTTGTCATCCGTGCAACTCCCGTCATTCCTGA 887

Qy   2492 AATTAAGGGGGCCTTCTTCAAATGTTTAGTGTGTGGTCACTCGCCTCCGCTAGTTACAGT 2551
Db    888 TATGACCACTGCTTTCTTCCGATGCTTGGTTTGTCAGCACACTGTTCAAGCGGATA---T 944

Qy   2552 TGTTAAAGGGCGGGTTGAGGAGCCAACAAGGTGTGAAAAGCCAGAATGTGCAGCACGGAA 2611
Db    945 CGACCGAGGAAGAATCAGCGAGCCAGAGCGATGTCCCCGAGACGTCTGTGGTTCAACTGG 1004

Qy   2612 TGCTATGTCTCTTATTCACAATCGATGCACTTTTGCAAATAAGCAGATAGTGCGTCTTCA 2671
Db   1005 CACTATGTCTCTTATTCATAACCGCTCCGAATTCACCGACAAACAAGTCATCCGTCTTCA 1064

Qy   2672 AGAAACTCCAGATGCCATTCCTGAAGGAGAGACTCCACACACAGTCAGCATGTGTTTATA 2731
Db   1065 GGAGACTCCCGATGCTGTACCCGATGGGCAGACTCCACATACTGTATCTCTTTGCGTCTA 1124

Qy   2732 CAACACTATGGTTGATGCTGTGAAGCCTGGAGATCGTATTGAGGTAACAGGAGTTTTCAA 2791
Db   1125 TGATGAACTCGTTGACTTGGTCAAGCCTGGTGACCGAGTGATCATCACTGGTATCTTCCG 1184

Qy   2792 GGCCATGGCAGTTCGAGTTGGTCCGAATCAACGAACATTACGAGCATTGTATAAGACCTA 2851
Db   1185 ATCAATCCCAGTTCGTGTCAACCCTCGCCAAAGGAGTATCAAATCTTTGTACAAGACCTA 1244

Qy   2852 CATCGATTGCGTGCACGTCAAGAAGTCTGACAGGGGTCGACTGCAAACTGAAGATCCTAT 2911
Db   1245 TCTCGATGTCGTCCACGTCAAGCGTACCAACACTGCCCGAATGGGCTTTGACCCTTCTA- 1303

Qy   2912 GGAGATGGATAAGGAGAATGATATGTATGCTGGGTATCATGAAAGTGATACTTCAGAAGC 2971
Db   1304 --------CTCGTGCGGGCGAGAGCAAGCCACCGGGTGACCTTGGGTCGCCTGTCAGATC 1355

Qy   2972 TGCTAATGAAGCAAAGATTCAAAAACTTAAAGAGCTGTCCAAGCTCCCGGGCATTTATGA 3031
Db   1356 AGCCGCTGCGGAGATGGAACAACGCATTATCGAGCTCTCCAACCACCCTGATCTTTACAA 1415

Qy   3032 TAGACTTTCAAGGTCGCTGGCTCCAAGCATTTGGGAGCTTGAAGATATTAAAAAGGGTCT 3091
Db   1416 CATACTTGCCTCATCTTTGGCGCCCTCCATCTACGAGCTCGAAGATGTTAAGAAGGGCAT 1475

Qy   3092 TCTTTGCCAGCTCTTTGGTGGGAAGGCTAAGAAAATTCCATCTGGAG------------- 3138
```

```
Db       1476 CCTCCTCCAGCTTTTCGGCGGTACCAACAAGTCTATCGCCCGTGGTGGCGGTGGGGGCGG 1535

Qy       3139 --CATCTTTCCGAGGTGACATCAATGTTTTACTTGTTGGGGACCCTGGTACCAGTAAATC 3196

Db       1536 TCCTAGATACAGAGGTGATATCAACGTGTTGATGGTTGGTGATCCTGGTACTAGTAAATC 1595

Qy       3197 TCAGCTGCTTCAGTATGTGCACAAGATAGCTCCTCGTGGAATCTACACTAGTGGGCGAGG 3256

Db       1596 TCAGATTCTGCAGTATGTGCATAAGATTGCTCCTCGAGGTGTCTACACCTCCGGTAAGGG 1655

Qy       3257 AAGTTCGGCGGTTGGGCTGACAGCGTATGTAACGAAGGATCCAGAAACTCGAGAGACGGT 3316

Db       1656 AAGTTCAGCGGTTGGTTTGACGGCCTATGTAACGAGAGATCCCGATTCCAAGCAGCTCGT 1715

Qy       3317 ATTGGAGAGCGGAGCTTTGGTTCTTAGTGATCGTGGGATATGCTGTATCGATGAGTTCGA 3376

Db       1716 TTTGGAAAGTGGTGCATTGGTTTTGTCTGACGGCGGAGTTTGCTGTATCGATGAGTTTGA 1775

Qy       3377 CAAAATGTCTGATAATGCCCGAAGCATGCTTCATGAGGTAATGGAGCAACAAACGGTATC 3436

Db       1776 CAAGATGAGTGATGCCACTCGAAGTGTCTTGCACGAAGTAATGGAGCAACAAACGGTTTC 1835

Qy       3437 TGTAGCCAAAGGGGGTATCATTGCCTCGCTGAACGCTCGGACGTCTGTCCTTGCATGTGC 3496

Db       1836 TATTGCCAAGGCAGGTATCATCACCACTCTCAATGCTCGTACATCCATTCTCGCCGCCGC 1895

Qy       3497 AAATCCTAGTGGGTCCCGATACAATGCGCGCCTTTCTGTGATTGATAACATCCAGCTTCC 3556

Db       1896 GAACCCAATCAACTCTCGATATGACCCTAACCTCCCTATTCCTGCCAACATCGATCTCCC 1955

Qy       3557 TCCAACTCTACTTTCTAGATTTGATTTAATTTACTTAATGCTCGACAAACCAGACGAGCA 3616

Db       1956 TCCTACATTGATCTCAAGATTCGACTTGCTGTACTTGGTGTTGGACAAGGTGGATGAGGT 2015

Qy       3617 AAACGATCGTCGTCTCGCCAGGCATCTCGTGGCTTTACACTATGAAAACTATGAAGTTTC 3676

Db       2016 GAACGACAGAAAGTTGGCAAAGCATTTGGTGGGACTCTATTTGTCAGATGTAGAGGATCA 2075

Qy       3677 AAAGCAGGAC---GCCTTAGATCTACAAACACTTACCGCGTATATCACCTATGCTCGTCA 3733

Db       2076 GCCTGCCGACAATATCATTCCTCTTCAGACCTTAACGTCCTACATTACATATGCCCGCTC 2135

Qy       3734 GCATGTACATCCTACATTAAGTGATGAAGCTGCTGAAGATTTGATTAATGGCTATGTTGA 3793

Db       2136 CAAAATCCACCCCGTCCTCACCGAAGGAGCTTCTGAAGCGCTTGTTCAGGCCTATGTTGA 2195

Qy       3794 GATGCGCCAAAAGGGCAACTTTCCTGGAAGCAGTAAAAAGGTGATAACAGCCACACCTCG 3853

Db       2196 GATGCGCAAGGCCGGTATGGACTCTCGTACTCAGGAAAAGAGGATCACTGCTACCACTAG 2255

Qy       3854 GCAACTCGAAAGTATGATTCGTATCAGTGAAGCCCTAGCTCGAATGAGATTTTCTGAAGT 3913

Db       2256 GCAGTTGGAAAGTATGATCAGATTGGGCGAAGCGCATGCGAGGATGAGATTAAGTGACAG 2315

Qy       3914 GGTAGAGAAAGTTGATGCAGCAGAAGCTGTGCGCCTTTTAGACGTCGCTTTGCAGCAATC 3973

Db       2316 GGTTGAGGAAGAGGACATTAGGGAGGCTGTTAGGTTGATCAAGAGTGCCCTGAGGGAGAG 2375

Qy       3974 TGCTACTGATCATGCAACAGGTACGATAGACATGGATCTTATCACGACTGGAGTGTCGGC 4033

Db       2376 TGCTACCGATCCCCTTACTGGTCAAATTGATCTTGACCTCATCAACACCGGTGCCGGCCA 2435

Qy       4034 CAGCGAGCGTATTCGTCGGGCCAACTTGCTAGCTGCTCTGCGAGAGCTTATAGCAGATAA 4093

Db       2436 AACCATGCGTCGCGCCCGAGCCGATCTTAAACGTGAAGTTATCAAACTCGTCGTCGAAAA 2495
```

RESULT 11
US-10-260-238-580
; Sequence 580, Application US/10260238
; Publication No. US20040016025A1
; GENERAL INFORMATION:
; APPLICANT: Budworth, Paul R.
; APPLICANT: Moughamer, Todd G.
; APPLICANT: Briggs, Steven P.
; APPLICANT: Cooper, Bret
; APPLICANT: Glazebrook, Jane
; APPLICANT: Goff, Stephen A.
; APPLICANT: Katagiri, Fumiyaki
; APPLICANT: Kreps, Joel
; APPLICANT: Provart, Nicholas
; APPLICANT: Ricke, Darrell
; APPLICANT: Zhu, Tong
; TITLE OF INVENTION: PROMOTERS FOR REGULATION OF PLANT EXPRESSION
; FILE REFERENCE: 60111-NP
; CURRENT APPLICATION NUMBER: US/10/260,238
; CURRENT FILING DATE: 2002-09-26
; PRIOR APPLICATION NUMBER: US 60/325,448
; PRIOR FILING DATE: 2001-09-26
; PRIOR APPLICATION NUMBER: US 60/325,277
; PRIOR FILING DATE: 2001-09-26
; PRIOR APPLICATION NUMBER: US 60/370,620
; PRIOR FILING DATE: 2002-04-04
; NUMBER OF SEQ ID NOS: 6077
; SEQ ID NO 580
; LENGTH: 1723
; TYPE: DNA
; ORGANISM: Oryza sativa
; FEATURE:
; NAME/KEY: N_region
; LOCATION: (95)..(95)
; OTHER INFORMATION: n = any nucleotide
US-10-260-238-580

```
  Query Match             11.3%;  Score 490;  DB 7;  Length 1723;
  Best Local Similarity   69.4%;  Pred. No. 2e-142;
  Matches  762;  Conservative    0;  Mismatches  285;  Indels   51;  Gaps    5;

Qy       2989 TTCAAAAACTTAAAGAGCTGTCCAAGCTCCCGGGCATTTATGATAGACTTTCAAGGTCGC 3048

Db        606 TTCACATAAAGAAGGAGTTGTCAAAGTTGCCAGATATATATGACAGATTAACTAGGTCAT 665

Qy       3049 TGGCTCCAAGCATTTGGGAGCTTGAAGATATTAAAAAGGGTCTTCTTTGC---CAGCTCT 3105

Db        666 TGGCTCCAAACATATGGGAGCTGGACGACGTTAAAAGGGGCCTCCTTTGCCAGCAGCTTT 725

Qy       3106 TTGGTGGGAAGGCTAAGAAAATTCCATCTGGAGCATCTTTCCGAGGTGACATCAATGTTT 3165

Db        726 TTGGTGGAAATGCTTTGAGGCTTCCTTCTGGAGCTAGTTTCCGAGGCGACATCAATATTT 785

Qy       3166 TACTTGTTGGGGACCCTGGTACCAGTAAATCTCAGCTGCTTCAGTATGTGCACAAGATAG 3225

Db        786 TGCTTGTTGGTGATCCTGGAACAAGTAAATCCCAGCTTCTCCAATACATGCACAAACTGT 845

Qy       3226 CTCCTCGTGGAATCTACACTAGTGGGCGAGGAAGTTCGGCGGTTGGGCTGACAGCGTATG 3285

Db        846 CTCCTCGTGGCATTTATACAAGTGGCAGAGGAAGTTCAGCTGTTGGCCTTACTGCATACG 905

Qy       3286 TAACGAAGGATCCAGAAACTCGAGAGACGGTATTGGAGAGCGGAGCTTTGGTTCTTAGTG 3345

Db        906 TTACCAAGGATCCTGAAACTGGTGAAACTGTTCTTGAGAGTGGAGCGCTTGTTTTGAGTG 965

Qy       3346 ATCGTGGGATATGCTGTATCGATGAGTTCGACAAAATGTCTGATAATGCCCGAAGCATGC 3405

Db        966 ACAAAGGTGTTTGTTGTATTGATGAATTTGATAAGATGTCTGATAATGCTCGAAGCATGT 1025

Qy       3406 TTCATGAGGTAATGGAGCAACAAACGGTATCTGTAGCCAAAGGGGGTATCATTGCCTCGC 3465

Db       1026 TACATGAGGTGATGGAACAACAGACTGTCTCCATTGCCAAGGCTGGAATAATTGCATCTT 1085

Qy       3466 TGAACGCTCGGACGTCTGTCCTTGCATGTGCAAATCCTAGTGGGTCCCGATACAATGCGC 3525

Db       1086 TAAATGCCAGAACATCAGTTCTAGCATGTGCAAATCCTACTGAATCACGTTATAATCCAA 1145

Qy       3526 GCCTTTCTGTGATTGATAACATCCAGCTTCCTCCAACTCTACTTTCTAGA--TTTGATTT 3583
```

```
Db        1146 GGCTCTCTGTGATTGACAATATCCATCTTCCTCCAACACTGCTTTCTAGAGGTTTGACCT 1205

Qy        3584 AATTTACTTAATGCTCGACAAACCAGACGAGCAAAACGATCGTCGTCTCGCCAGGCATCT 3643
Db        1206 CATTTATCTGATATTGGACAAGGCAGATGAGCAAACTGATAGACGCCTGGCTAAGCATAT 1265

Qy        3644 CGTGGCTTTACACTATGAAAACTATGAAGTTTCAAAGCAGGACGCCTTAGATCTACAAAC 3703
Db        1266 TGTTTCGTTGCATTTTGAGAA------------------------------------- 1286

Qy        3704 ACTTACCGCGTATATCACCTATGCTCGTCAGCATGTACATCCTACATTAAGTGATGAAGC 3763
Db        1287 ---TCCAAACTACATAAGTTATGCAAGGAAGCATATACAACCACAGTTATCTGATGAAGC 1343

Qy        3764 TGCTGAAGATTTGATTAATGGCTATGTTGAGATGCGCCAAAAGGGCAACTTTCCTGGAAG 3823
Db        1344 TGCAGAAGAATTGACTCGCGGCTATGTTGAGATGAGGAAAAGAGGAAACAGCCCTGGTAG 1403

Qy        3824 CAGTAAAAAGGTGATAACAGCCACACCTCGGCAACTCGAAAGTATGATTCGTATCAGTGA 3883
Db        1404 CAG---AAAGAAGATAACTGCGACAGCTCGACAAATTGAGAGCTTGATTCGGCTCAGTGA 1460

Qy        3884 AGCCCTAGCTCGAATGAGATTTTCTGAAGTGGTAGAGAAAGTTGATGCAGCAGAAGCTGT 3943
Db        1461 AGCACTGGCCCGAATGCGATTCTCTGAAATGGTTGAAGTACAAGATGTTGTAGAGGCCTT 1520

Qy        3944 GCGCCTTTTAGACGTCGCTTTGCAGCAATCTGCTACTGATCATGCAAC-AGGTACGATAG 4002
Db        1521 CAGGCTTCTCGAAGTTGCCATGCAGCAATCGGCAACTGATCATGCCACTGGGTACAATCG 1580

Qy        4003 ACATGGATCTTATCACGACTGGAGTGTCGGCCAGCGAGCGTATTCGTCGGGCCAACTTGC 4062
Db        1581 ATATGGATCTTATCATGACTGGAATATCTGCGAGCGAAAGGCAGAGGCGGGACCAGTTGC 1640

Qy        4063 TAGCTGCTCTGCGAGAGC 4080
Db        1641 TGGAAGAAATTAGGAAGC 1658
```

RESULT 12
US-10-032-585-6480
; Sequence 6480, Application US/10032585
; Publication No. US20030180953A1
; GENERAL INFORMATION:
; APPLICANT: Terry, Roemer D.
; APPLICANT: Bo, Jiang
; APPLICANT: Charles, Boone
; APPLICANT: Howard, Bussey
; TITLE OF INVENTION: Gene Disruption Methodologies for Drug Target Discovery
; FILE REFERENCE: 10182-005-999
; CURRENT APPLICATION NUMBER: US/10/032,585
; CURRENT FILING DATE: 2001-12-20
; NUMBER OF SEQ ID NOS: 8000
; SOFTWARE: PatentIn version 3.1
; SEQ ID NO 6480
; LENGTH: 2733
; TYPE: DNA
; ORGANISM: Candida albicans
US-10-032-585-6480

```
  Query Match             11.0%;  Score 480.2;  DB 6;  Length 2733;
  Best Local Similarity   56.6%;  Pred. No. 3.4e-139;
  Matches  959;  Conservative    0;  Mismatches  718;  Indels   18;  Gaps    3;

Qy        2409 ATGCGTGAACTCAACCCTTCAGATATAGACAAATTGGTTTCTGTTAAAGGAATGGTTATC 2468
Db         883 ATTCGTGAATTGAATCCTAATGATATTGATAAGTTAGTTAGTGTTAAAGGGTTAACATTA 942

Qy        2469 CGGTGCAGTTCTATCATACCTGAAATTAAGGGGGCCTTCTTCAAATGTTTAGTGTGTGGT 2528
Db         943 CGTTCTACTTCTATCATTCCTGATATGAAAGTGGCATTTTTCAGATGTAACGCCTGTGGT 1002

Qy        2529 CACTCGCCTCCGCTAGTTACAGTTGTTAAAGGGCGGGTTGAGGAGCCAACAAGGTGTGAA 2588
Db        1003 CATACTGTTGGG---GTGGAAATTGATCGTGGTGTAATTTCCGAACCAACAAAATGTCCT 1059

Qy        2589 AAGCCAGAATGTGCAGCACGGAATGCTATGTCTCTTATTCACAATCGATGCACTTTTGCA 2648
Db        1060 AGAGAAGTATGTGGTCAAACCAATTCAATGGTACTTATTCATAATCGTTCATCATTTTCT 1119

Qy        2649 AATAAGCAGATAGTGCGTCTTCAAGAAACTCCAGATGCCATTCCTGAAGGAGAGACTCCA 2708
Db        1120 GATAAACAAGTGATTAAATTACAAGAAACTCCTGATTTGGTACCTGATGGTCAAACCCCT 1179

Qy        2709 CACACAGTCAGCATGTGTTTATACAACACTATGGTTGATGCTGTGAAGCCTGGAGATCGT 2768
Db        1180 CATTCTATCAATTTATGTGTTTACGATGAATTGGTTGATTCTTGTCGTGCTGGTGATCGT 1239

Qy        2769 ATTGAGGTAACAGGAGTTTTCAAGGCCATGGCAGTTCGAGTTGGTCCGAATCAACGAACA 2828
Db        1240 GTTGAAGTATGTGGTATTTTCAGATCAACTCCAGTTCGAGCAAATCCTAGACAAAGAGCT 1299

Qy        2829 TTACGAGCATTGTATAAGACCTACATCGATTGCGTGCACGTCAAGAAGTCTGACAGGGGT 2888
Db        1300 TTAAAAAACTTATATAAAACATATTTGGATATTGTTCATGTGAAAAAAATCGATAAAAGG 1359

Qy        2889 CGACTGCAAACTGAAGATCCTATGGAGATGGATAAGGAGAATGATATGTATGCTGGGTAT 2948
Db        1360 CGATTAGGTGGTGATGTTACCA---------CTTTAGAACACGAATTGGCTGAAAAGGAT 1410

Qy        2949 CATGAAAGTGATACTTCAGAAGCTGCTAATGAAGCAAAGATTCAAAAACTTAAAGAGCTG 3008
Db        1411 CAAGAAGTTGAACAAGTTCGAAAAATCACTGCTGAAGAAGAAGCTAAAATCAAAGAAATT 1470

Qy        3009 TCCAAGCTCCCGGGCATTTATGATAGACTTTCAAGGTCGCTGGCTCCAAGCATTTGGGAG 3068
Db        1471 TCTGAACGTGATGATTTGTATGAAATATTGGCTAGATCATTAGCTCCATCTATTTATGAG 1530

Qy        3069 CTTGAAGATATTAAAAAGGGTCTTCTTTGCCAGCTCTTTGGTGGGAAGGCTAAGAAAATT 3128
Db        1531 ATGGATGATGTTAAAAAGGGGATCTTGTTACAATTATTTGGTGGTACTAATAAAACATTT 1590

Qy        3129 CCATCTGGAGCATCTTTCCGAGGTGACATCAATGTTTTACTTGTTGGGGACCCTGGTACC 3188
Db        1591 ACAAAAGGTGGTCGTTATAGAGGTGATATAAATATTTTATTATGTGGTGATCCATCTACT 1650

Qy        3189 AGTAAATCTCAGCTGCTTCAGTATGTGCACAAGATAGCTCCTCGTGGAATCTACACTAGT 3248
Db        1651 TCTAAATCACAAATCTTACAATATGTTCATAAGATTGCTCCAAGAGGGGTTTATACTTCT 1710

Qy        3249 GGGCGAGGAAGTTCGGCGGTTGGGCTGACAGCGTATGTAACGAAGGATCCAGAAACTCGA 3308
Db        1711 GGTAAAGGTTCTTCGGCTGTTGGTTTGACAGCTTATATCACCAGAGATATTGATACAAAA 1770

Qy        3309 GAGACGGTATTGGAGAGCGGAGCTTTGGTTCTTAGTGATCGTGGGATATGCTGTATCGAT 3368
Db        1771 CAATTAGTATTGGAAAGTGGGGCCCTTGTGTTATCTGATGGAGGTGTTTGTTGTATTGAT 1830

Qy        3369 GAGTTCGACAAAATGTCTGATAATGCCCGAAGCATGCTTCATGAGGTAATGGAGCAACAA 3428
Db        1831 GAGTTTGACAAGATGAGTGATGCCACCAGATCAGTTTTACATGAAGTCATGGAACAACAA 1890

Qy        3429 ACGGTATCTGTAGCCAAAGGGGGTATCATTGCCTCGCTGAACGCTCGGACGTCTGTCCTT 3488
Db        1891 ACTATTTCTATTGCTAAAGCCGGTATTATTACCACATTAAATGCTAGGACATCTATTTTA 1950

Qy        3489 GCATGTGCAAATCCTAGTGGGTCCCGATACAATGCGCGCCTTTCTGTGATTGATAACATC 3548
Db        1951 GCTTCTGCAAATCCAATTAATTCTCGTTATGATCCAAATTTGCCTGTGACGGGTAACATT 2010

Qy        3549 CAGCTTCCTCCAACTCTACTTTCTAGATTTGATTTAATTTACTTAATGCTCGACAAACCA 3608
Db        2011 GATTTGCCACCACCATTGTTATCACGTTTTGATTTGGTTTATTTGATTTTGGATAAAGTT 2070

Qy        3609 GACGAGCAAAACGATCGTCGTCTCGCCAGGCATCTCGTGGCTTTACACTATGAAAACTAT 3668
```

```
Db       2071 GATGAATCAATCGATAGACAATTGGCTAGACATTTGACTGATATGTATTTGGAAGATGCT 2130

Qy       3669 GAAGTTTCAAAGCAGGACGCCTTAGATCTAC------AAACACTTACCGCGTATATCACC 3722

Db       2131 CCTGAAACTGTTAATGCCAATTCGGTATTACCCGTGGAATTATTAACACTTTATATTCAA 2190

Qy       3723 TATGCTCGTCAGCATGTACATCCTACATTAAGTGATGAAGCTGCTGAAGATTTGATTAAT 3782

Db       2191 TATGCTAAAGAGAATTTCAATCCCGTGATGACAGAAGAAGGTAAAAATGAATTAGTTAGA 2250

Qy       3783 GGCTATGTTGAGATGCGCCAAAAGGGCAACTTTCCTGGAAGCAGTAAAAAGGTGATAACA 3842

Db       2251 TCATACGTTGAAATGAGAAAATTGGGGGAGGATGCTAGATCATCAGAAAAGAGAATCACC 2310

Qy       3843 GCCACACCTCGGCAACTCGAAAGTATGATTCGTATCAGTGAAGCCCTAGCTCGAATGAGA 3902

Db       2311 GCTACTACAAGACAATTGGAATCGATGATTAGATTGTCTGAAGCTCATGCTAAAATGAGA 2370

Qy       3903 TTTTCTGAAGTGGTAGAGAAAGTTGATGCAGCAGAAGCTGTGCGCCTTTTAGACGTCGCT 3962

Db       2371 TTATCTGAAAGAGTTGAATTGATCGATGTTAAAGAAGCAGTAAGATTAATTAAATCGGCC 2430

Qy       3963 TTGCAGCAATCTGCTACTGATCATGCAACAGGTACGATAGACATGGATCTTATCACGACT 4022

Db       2431 ATTAAAGATTATGCTACTGATCCAGTTACTGGTAGAATCGATATGGATATGATTCAAACT 2490

Qy       4023 GGAGTGTCGGCCAGCGAGCGTATTCGTCGGGCCAACTTGCTAGCTGCTCTGCGAGAGCTT 4082

Db       2491 GGTACTACGGCACAACAAAGAAGAGTACAAGAAGATTTGGTCAGTGAAATTATGAAAATC 2550

Qy       4083 ATAGCAGATAAAATT 4097

Db       2551 ATCGAAGAAAATAAT 2565
```

RESULT 13
US-10-369-493-46198
; Sequence 46198, Application US/10369493
; Publication No. US20030233675A1
; GENERAL INFORMATION:
; APPLICANT: Cao, Yongwei
; APPLICANT: Hinkle, Gregory J.
; APPLICANT: Slater, Steven C.
; APPLICANT: Goldman, Barry S.
; APPLICANT: Chen, Xianfeng
; TITLE OF INVENTION: EXPRESSION OF MICROBIAL PROTEINS IN PLANTS FOR PRODUCTION OF
; TITLE OF INVENTION: PLANTS WITH IMPROVED PROPERTIES
; FILE REFERENCE: 38-10(52052)B
; CURRENT APPLICATION NUMBER: US/10/369,493
; CURRENT FILING DATE: 2003-02-28
; PRIOR APPLICATION NUMBER: US 60/360,039
; PRIOR FILING DATE: 2002-02-21
; NUMBER OF SEQ ID NOS: 47374
; SEQ ID NO 46198
; LENGTH: 2802
; TYPE: DNA
; ORGANISM: Saccharomyces cerevisiae
US-10-369-493-46198

```
  Query Match             10.8%;  Score 470.8;  DB 6;  Length 2802;
  Best Local Similarity   55.6%;  Pred. No. 3.2e-136;
  Matches  948;  Conservative    0;  Mismatches  747;  Indels    9;  Gaps    2;

Qy       2364 AAGCATATTGAGGCCAGACCTTTCAATCTCAAAGCATCGGTGCACATGCGTGAACTCAAC 2423

Db        895 AAGTTTTATAAAGTAAGGCCTTACAATGTAGGTTCTTGTAAAGGTATGCGCGAATTGAAT 954

Qy       2424 CCTTCAGATATAGACAAATTGGTTTCTGTTAAAGGAATGGTTATCCGGTGCAGTTCTATC 2483

Db        955 CCAAATGATATTGATAAATTGATAAACTTAAAGGGTCTTGTCCTGAGATCAACTCCGGTT 1014

Qy       2484 ATACCTGAAATTAAGGGGGCCTTCTTCAAATGTTTAGTGTGTGGTCACTCGCCTCCGCTA 2543

Db       1015 ATCCCTGATATGAAGGTAGCGTTTTTCAAATGCAACGTCTGCGATCA---TACAATGGCA 1071

Qy       2544 GTTACAGTTGTTAAAGGGCGGGTTGAGGAGCCAACAAGGTGTGAAAAGCCAGAATGTGCA 2603

Db       1072 GTGGAAATTGATAGAGGAGTTATACAAGAGCCCGCTAGGTGTGAACGTATTGATTGTAAT 1131

Qy       2604 GCACGGAATGCTATGTCTCTTATTCACAATCGATGCACTTTTGCAAATAAGCAGATAGTG 2663

Db       1132 GAACCCAATTCCATGTCACTGATTCACAACAGGTGTTCATTCGCAGATAAACAAGTCATT 1191

Qy       2664 CGTCTTCAAGAAACTCCAGATGCCATTCCTGAAGGAGAGACTCCACACACAGTCAGCATG 2723

Db       1192 AAGTTACAGGAAACTCCAGACTTTGTGCCTGATGGACAAACGCCTCACTCTATCTCATTA 1251

Qy       2724 TGTTTATACAACACTATGGTTGATGCTGTGAAGCCTGGAGATCGTATTGAGGTAACAGGA 2783

Db       1252 TGTGTTTACGATGAATTAGTGGATTCTTGCAGGGCGGGCGATCGTATTGAGGTGACTGGC 1311

Qy       2784 GTTTTCAAGGCCATGGCAGTTCGAGTTGGTCCGAATCAACGAACATTACGAGCATTGTAT 2843

Db       1312 ACGTTCAGGTCCATCCCCATTAGAGCTAATTCCAGGCAACGCGTACTAAAGTCGTTGTAT 1371

Qy       2844 AAGACCTACATCGATTGCGTGCACGTCAAGAAGTCTGACAGGGGTCGACTGCAAACTGAA 2903

Db       1372 AAAACATACGTCGATGTGGTCCACGTTAAAAAAGTTTCAGATAAGAGGTTAGACGTCGAT 1431

Qy       2904 GATCCTATGGAGATGGATAAGGAGAATGATATGTATGCTGGGTATCATGAAAGTGATACT 2963

Db       1432 ACTTCTACTATTGAACAAGAATTAATGCAGAACAAGGTAGATCATAACGAGGTCGAAGAA 1491

Qy       2964 TCAGAAGCTGCTAATGAAGCAAAGATTCAAAAACTTAAAGAGCTGTCCAAGCTCCCGGGC 3023

Db       1492 GTAAGACAAATTACTGATCAGGATTTAGCAAAAATTCGCGAGGTTGCGGCAAGAGAAGAT 1551

Qy       3024 ATTTATGATAGACTTTCAAGGTCGCTGGCTCCAAGCATTTGGGAGCTTGAAGATATTAAA 3083

Db       1552 TTATACAGTTTATTAGCCCGCTCTATTGCCCCAAGTATTTACGAGCTAGAAGATGTCAAG 1611

Qy       3084 AAGGGTCTTCTTTGCCAGCTCTTTGGTGGGAAGGCTAAGAAAATTCCATCTGGAGCATCT 3143

Db       1612 AAGGGTATATTACTTCAGCTATTTGGCGGCACGAATAAAACCTTTACCAAGGGTGGCCGT 1671

Qy       3144 TTCCGAGGTGACATCAATGTTTTACTTGTTGGGGACCCTGGTACCAGTAAATCTCAGCTG 3203

Db       1672 TATAGAGGTGACATAAATATTTTACTTTGTGGGGATCCTTCTACTTCCAAATCGCAAATT 1731

Qy       3204 CTTCAGTATGTGCACAAGATAGCTCCTCGTGGAATCTACACTAGTGGGCGAGGAAGTTCG 3263

Db       1732 TTGCAATACGTTCACAAAATTACTCCTCGTGGTGTATACTTCGGGTAAAGGTTCATCT 1791

Qy       3264 GCGGTTGGGCTGACAGCGTATGTAACGAAGGATCCAGAAACTCGAGAGACGGTATTGGAG 3323

Db       1792 GCCGTTGGTTTAACTGCTTATATTACAAGGGATGTCGACACAAAACAACTTGTTTGGAA 1851

Qy       3324 AGCGGAGCTTTGGTTCTTAGTGATCGTGGGATATGCTGTATCGATGAGTTCGACAAAATG 3383

Db       1852 AGTGGTGCATTAGTATTGTCTGATGGAGGTGTTTGTTGTATTGACGAGTTTGATAAAATG 1911

Qy       3384 TCTGATAATGCCCGAAGCATGCTTCATGAGGTAATGGAGCAACAAACGGTATCTGTAGCC 3443

Db       1912 AGTGATTCTACAAGGTCCGTCTTGCACGAAGTCATGGAACAGCAGACTATTTCAATCGCA 1971

Qy       3444 AAAGGGGGTATCATTGCCTCGCTGAACGCTCGGACGTCTGTCCTTGCATGTGCAAATCCT 3503

Db       1972 AAAGCGGGAATTATCACAACACTTAATGCCAGAAGTTCTATTTTGGCCAGTGCTAACCCA 2031

Qy       3504 AGTGGGTCCCGATACAATGCGCGCCTTTCTGTGATTGATAACATCCAGCTTCCTCCAACT 3563

Db       2032 ATTGGTTCACGCTACAACCCCAATTTGCCTGTGACTGAAAATATTGATCTACCGCCCCCA 2091

Qy       3564 CTACTTTCTAGATTTGATTTAATTTACTTAATGCTCGACAAACCAGACGAGCAAAACGAT 3623

Db       2092 CTACTTTCGAGATTCGATCTGGTCTATCTTGTCCTTGATAAGGTTGATGAGAAAAATGAC 2151
```

```
Qy     3624 CGTCGTCTCGCCAGGCATCTCGTGGCTTTACACTATGAAAA------CTATGAAGTTTCA 3677
Db     2152 AGAGAACTAGCCAAACACTTAACAAATCTTTACCTGGAAGATAAGCCCGAACATATTTCT 2211
Qy     3678 AAGCAGGACGCCTTAGATCTACAAACACTTACCGCGTATATCACCTATGCTCGTCAGCAT 3737
Db     2212 CAAGACGACGTGCTACCAGTTGAATTTTTAACGATGTATATTAGTTATGCAAAGGAGCAC 2271
Qy     3738 GTACATCCTACATTAAGTGATGAAGCTGCTGAAGATTTGATTAATGGCTATGTTGAGATG 3797
Db     2272 ATACACCCAATAATCACCGAGGCCGCTAAGACTGAGCTTGTTCGTGCTTATGTAGGAATG 2331
Qy     3798 CGCCAAAAGGGCAACTTTCCTGGAAGCAGTAAAAAGGTGATAACAGCCACACCTCGGCAA 3857
Db     2332 AGAAAGATGGGTGACGATTCGAGATCCGATGAGAAGAGAATCACAGCTACCACAAGACAA 2391
Qy     3858 CTCGAAAGTATGATTCGTATCAGTGAAGCCCTAGCTCGAATGAGATTTTCTGAAGTGGTA 3917
Db     2392 CTTGAAAGTATGATTCGTTTGGCTGAGGCGCACGCCAAAATGAAATTGAAAAACGTCGTA 2451
Qy     3918 GAGAAAGTTGATGCAGCAGAAGCTGTGCGCCTTTTAGACGTCGCTTTGCAGCAATCTGCT 3977
Db     2452 GAGCTGGAGGATGTTCAAGAAGCCGTTAGATTAATTAGATCAGCCATAAAAGATTATGCA 2511
Qy     3978 ACTGATCATGCAACAGGTACGATAGACATGGATCTTATCACGACTGGAGTGTCGGCCAGC 4037
Db     2512 ACAGACCCTAAAACCGGTAAAATCGACATGAATTTAGTTCAAACAGGTAAATCAGTTATT 2571
Qy     4038 GAGCGTATTCGTCGGGCCAACTTG 4061
Db     2572 CAGAGAAAACTACAGGAGGATTTG 2595
```

RESULT 14
US-10-128-714-2558
; Sequence 2558, Application US/10128714
; Publication No. US20030119013A1
; GENERAL INFORMATION:
; APPLICANT: Jiang, Bo
; APPLICANT: Hu, Wenqi
; APPLICANT: Tishkoff, Daniel
; APPLICANT: Zamudio, Carlos
; APPLICANT: Eroshkin, Alexey M
; APPLICANT: Lemieux, Sebastien M
; TITLE OF INVENTION: Identification of Essential Genes in Aspergillus fumigatus and
; TITLE OF INVENTION: Methods of Use
; FILE REFERENCE: 10182-018-999
; CURRENT APPLICATION NUMBER: US/10/128,714
; CURRENT FILING DATE: 2002-04-23
; PRIOR APPLICATION NUMBER: US 60/285,697
; PRIOR FILING DATE: 2001-04-23
; PRIOR APPLICATION NUMBER: US 60/287,066
; PRIOR FILING DATE: 2001-04-27
; PRIOR APPLICATION NUMBER: US 60/295,890
; PRIOR FILING DATE: 2001-06-05
; PRIOR APPLICATION NUMBER: US 60/303,899
; PRIOR FILING DATE: 2001-07-09
; PRIOR APPLICATION NUMBER: US 60/316,362
; PRIOR FILING DATE: 2001-08-31
; NUMBER OF SEQ ID NOS: 8603
; SOFTWARE: PatentIn version 3.1
; SEQ ID NO 2558
; LENGTH: 2625
; TYPE: DNA
; ORGANISM: Aspergillus fumigatus
US-10-128-714-2558

```
  Query Match             10.4%;  Score 451.6;  DB 5;  Length 2625;
  Best Local Similarity   56.1%;  Pred. No. 3.4e-130;
  Matches  944;  Conservative    0;  Mismatches  719;  Indels   21;  Gaps    4;

Qy     2382 CCTTTCAATCTCAAAGCATCGGTGCACATGCGTGAACTCAACCCTTCAGATATAGACAAA 2441
Db      724 CCTTTTGGTCTCGACTCGACTGTGAATATGAGAGACCTCGATCCTGCAGACATGGATAAA 783
Qy     2442 TTGGTTTCTGTTAAAGGAATGGTTATCCGGTGCAGTTCTATCATACCTGAAATTAAGGGG 2501
Db      784 CTAGTAAGCATTAAGGGCTTAGTCATTCGAACGACACCCATCATTCCTGATATGAAAGAA 843
Qy     2502 GCCTTCTTCAAATGTTTAGTGTGTGGTCACTCGCCTCCGCTAGTTACAGTTGTTAAAGGG 2561
Db      844 GCTTTCTTCCGTTGCCAAGTCTGCAACCATGGTGTTCAGGTCGACA---TTGATCGTGGA 900
Qy     2562 CGGGTTGAGGAGCCAACAAGGTGTGAAAAGCCAGAATGTGCAGCACGGAATGCTATGTCT 2621
Db      901 AAGATTGCGGAACCCACTGAGTGCCCACGTCCAGTGTGTAAGGAACGAAACTCGATGCAA 960
Qy     2622 CTTATTCACAATCGATGCACTTTTGCAAATAAGCAGATAGTGCGTCTTCAAGAAACTCCA 2681
Db      961 CTCATCCATAACCGCTGTGTATTTGCCGACAAGCAGGTCATCAAGTTGCAGGAAACACCT 1020
Qy     2682 GATGCCATTCCTGAAGGAGAGACTCCACACACAGTCAGCATGTGTTTATACAACACTATG 2741
Db     1021 GACAGCATTCCTGATGGCCAGACTCCTCACTCGGTTTCCCTTTGTGTGTATGATGAGCTG 1080
Qy     2742 GTTGATGCTGTGAAGCCTGGAGATCGTATTGAGGTAACAGGAGTTTTCAAGGCCATGGCA 2801
Db     1081 GTGGATGTCTGCAAGGCTGGTGATCGGGTCGAAGTGACCGGTATTTTCCGGTGCAACCCT 1140
Qy     2802 GTTCGAGTTGGTCCGAATCAACGAACATTACGAGCATTGTATAAGACCTACATCGATTGC 2861
Db     1141 GTGCGCGTTAATCCTCGCCAGCGTACACAGAAGTCGCTGTTCAAGACGTACATAGATGTT 1200
Qy     2862 GTGCACGTCAAGAAGTCTGACAGGGGTCGACTGCAAACTGAAGATCCTATGGAGATGGAT 2921
Db     1201 CTTCATGTTCAGAAGATCGATCGCAAGAAGTTGGGTATCGACGTCTCGACCATCGAGCAG 1260
Qy     2922 AAGGAGAATGATATGTATGCTGGGTATCATGAAAGTGATACTTCAGAAGCTGCTAATGAA 2981
Db     1261 GAGCTCTCGGAACAGGCGGCTGGGGATGCAGAACAAACACGTAGGCTCACTGCGGAGGAG 1320
Qy     2982 GCAAAGATTCAAAAACTTAAAGAGCTGTCCAAGCTCCCGGGCATTTATGATAGACTTTCA 3041
Db     1321 GAAGAG------AAAATTAAGCGAACTGCTACCAGACCTGATCTGTATGAGCTTCTCTCT 1374
Qy     3042 AGGTCGCTGGCTCCAAGCATTTGGGAGCTTGAAGATATTAAAAAGGGTCTTCTTTGCCAG 3101
Db     1375 CGGTCCTTGGCCCCCAGCATCTACGAGATGGACGACGTGAAGAAGGGAATCCTGCTTCAG 1434
Qy     3102 CTCTTTGGTGGGAAGGCTAAGAAAATTCCATCTGGAG------CATCTTTCCGAGGTGAC 3155
Db     1435 TTGTTTGGAGGCACCAACAAGACCTTCCAGAAGGGTGGTAACCCACGATACCGTGGAGAT 1494
Qy     3156 ATCAATGTTTTACTTGTTGGGGACCCTGGTACCAGTAAATCTCAGCTGCTTCAGTATGTG 3215
Db     1495 ATCAATATCCTTCTCTGTGGTGACCCATCTACATCCAAGTCCCAGCTTCTTCGTTACGTC 1554
Qy     3216 CACAAGATAGCTCCTCGTGGAATCTACACTAGTGGGCGAGGAAGTTCGGCGGTTGGGCTG 3275
Db     1555 CATAAGATTGCCCCTCGCGGTGTGTATACCAGCGGCAAGGGCTCCTCGGCTGTTGGTCTT 1614
Qy     3276 ACAGCGTATGTAACGAAGGATCCAGAAACTCGAGAGACGGTATTGGAGAGCGGAGCTTTG 3335
Db     1615 ACGGCGTACGTCACCCGCGATCCTGAAACCCGCCAGATGGTCCTCGAGTCGGGTGCCTTG 1674
Qy     3336 GTTCTTAGTGATCGTGGGATATGCTGTATCGATGAGTTCGACAAAATGTCTGATAATGCC 3395
Db     1675 GTTCTTTCAGACGGCGGTATCTGTTGCATCGACGAGTTCGACAAGATGAACGAATCCACT 1734
Qy     3396 CGAAGCATGCTTCATGAGGTAATGGAGCAACAAACGGTATCTGTAGCCAAAGGGGGTATC 3455
Db     1735 CGGTCCGTTCTGCATGAAGTCATGGAACAACAGACAGTATCTATCGCCAAGGCAGGCATT 1794
Qy     3456 ATTGCCTCGCTGAACGCTCGGACGTCTGTCCTTGCATGTGCAAATCCTAGTGGGTCCCGA 3515
```

```
Db     1795 ATCACTACTTTGAACGCTAGGACCAGCATCCTGGCTTCCGCCAATCCGATCGGTAGCAGG 1854

Qy     3516 TACAATGCGCGCCTTTCTGTGATTGATAACATCCAGCTTCCTCCAACTCTACTTTCTAGA 3575

Db     1855 TACAATCCCAACTTGCCCGTTCCTCAAAATATTGACCTTCCGCCTACCTTGCTCTCCCGA 1914

Qy     3576 TTTGATTTAATTTACTTAATGCTCGACAAACCAGACGAGCAAAACGATCGTCGTCTCGCC 3635

Db     1915 TTCGACTTGGTATACCTCGTGCTGGACCGAGTGGATGAGCAGGAAGATCGTCGGCTCGCT 1974

Qy     3636 AGGCATCTCGTGGCTTTACACTATGAAAAC------TATGAAGTTTCAAAGCAGGACGCC 3689

Db     1975 AAGCATCTTGTCAATATGTACCTGGAAGACAGACCTGAGCATGCTGCCGAGCAAGAAATC 2034

Qy     3690 TTAGATCTACAAACACTTACCGCGTATATCACCTATGCTCGTCAGCATGTACATCCTACA 3749

Db     2035 TTGCCGATCGAATTCCTTACAGCCTATATCACCTACGCCAAGACCAAAGTCCATCCAGTG 2094

Qy     3750 TTAAGTGATGAAGCTGCTGAAGATTTGATTAATGGCTATGTTGAGATGCGCCAAAAGGGC 3809

Db     2095 CTCACACCGGCCGCCGGTAAAGCCTTGTCGGATGCTTACGTTAACATGCGTAAGCTTGGA 2154

Qy     3810 AACTTTCCTGGAAGCAGTAAAAAGGTGATAACAGCCACACCTCGGCAACTCGAAAGTATG 3869

Db     2155 GATGACATCCGGTCTTCTGACCGCCGTATCACCGCTACCACTCGTCAACTGGAGTCCATG 2214

Qy     3870 ATTCGTATCAGTGAAGCCCTAGCTCGAATGAGATTTTCTGAAGTGGTAGAGAAAGTTGAT 3929

Db     2215 ATCCGACTGTCGGAAGCGCATGCGCGTATGCGGCTATCGCCGGAGGTCACTGCGGATGAT 2274

Qy     3930 GCAGCAGAAGCTGTGCGCCTTTTAGACGTCGCTTTGCAGCAATCTGCTACTGATCATGCA 3989

Db     2275 GTGGAGGAAGCCGTGCGCCTGATCCGCTCCGCCATCAAGCAGGCGGCCACTGACTCTCGG 2334

Qy     3990 ACAGGTACGATAGACATGGATCTTATCACGACTGGAGTGTCGGCCAGCGAGCGTATTCGT 4049

Db     2335 ACCGGTCTGATCGACATGAGCTTGTTGACGGAGGGCACTAGTGCCAGCGAGAGACGCAGC 2394

Qy     4050 CGGG 4053

Db     2395 CGGG 2398
```

RESULT 15
US-10-128-714-7558
; Sequence 7558, Application US/10128714
; Publication No. US20030119013A1
; GENERAL INFORMATION:
; APPLICANT: Jiang, Bo
; APPLICANT: Hu, Wenqi
; APPLICANT: Tishkoff, Daniel
; APPLICANT: Zamudio, Carlos
; APPLICANT: Eroshkin, Alexey M
; APPLICANT: Lemieux, Sebastien M
; TITLE OF INVENTION: Identification of Essential Genes in Aspergillus fumigatus and
; TITLE OF INVENTION: Methods of Use
; FILE REFERENCE: 10182-018-999
; CURRENT APPLICATION NUMBER: US/10/128,714
; CURRENT FILING DATE: 2002-04-23
; PRIOR APPLICATION NUMBER: US 60/285,697
; PRIOR FILING DATE: 2001-04-23
; PRIOR APPLICATION NUMBER: US 60/287,066
; PRIOR FILING DATE: 2001-04-27
; PRIOR APPLICATION NUMBER: US 60/295,890
; PRIOR FILING DATE: 2001-06-05
; PRIOR APPLICATION NUMBER: US 60/303,899
; PRIOR FILING DATE: 2001-07-09
; PRIOR APPLICATION NUMBER: US 60/316,362
; PRIOR FILING DATE: 2001-08-31
; NUMBER OF SEQ ID NOS: 8603
; SOFTWARE: PatentIn version 3.1
; SEQ ID NO 7558
;   LENGTH: 3072
;   TYPE: DNA
;   ORGANISM: Aspergillus fumigatus
US-10-128-714-7558

```
  Query Match             10.4%;  Score 451.6;  DB 5;  Length 3072;
  Best Local Similarity   56.1%;  Pred. No. 3.8e-130;
  Matches  944;  Conservative    0;  Mismatches  719;  Indels   21;  Gaps    4;

Qy     2382 CCTTTCAATCTCAAAGCATCGGTGCACATGCGTGAACTCAACCCTTCAGATATAGACAAA 2441

Db     1171 CCTTTTGGTCTCGACTCGACTGTGAATATGAGAGACCTCGATCCTGCAGACATGGATAAA 1230

Qy     2442 TTGGTTTCTGTTAAAGGAATGGTTATCCGGTGCAGTTCTATCATACCTGAAATTAAGGGG 2501

Db     1231 CTAGTAAGCATTAAGGGCTTAGTCATTCGAACGACACCCATCATTCCTGATATGAAAGAA 1290

Qy     2502 GCCTTCTTCAAATGTTTAGTGTGTGGTCACTCGCCTCCGCTAGTTACAGTTGTTAAAGGG 2561

Db     1291 GCTTTCTTCCGTTGCCAAGTCTGCAACCATGGTGTTCAGGTCGACA---TTGATCGTGGA 1347

Qy     2562 CGGGTTGAGGAGCCAACAAGGTGTGAAAAGCCAGAATGTGCAGCACGGAATGCTATGTCT 2621

Db     1348 AAGATTGCGGAACCCACTGAGTGCCCACGTCCAGTGTGTAAGGAACGAAACTCGATGCAA 1407

Qy     2622 CTTATTCACAATCGATGCACTTTTGCAAATAAGCAGATAGTGCGTCTTCAAGAAACTCCA 2681

Db     1408 CTCATCCATAACCGCTGTGTATTTGCCGACAAGCAGGTCATCAAGTTGCAGGAAACACCT 1467

Qy     2682 GATGCCATTCCTGAAGGAGAGACTCCACACACAGTCAGCATGTGTTTATACAACACTATG 2741

Db     1468 GACAGCATTCCTGATGGCCAGACTCCTCACTCGGTTTCCCTTTGTGTGTATGATGAGCTG 1527

Qy     2742 GTTGATGCTGTGAAGCCTGGAGATCGTATTGAGGTAACAGGAGTTTTCAAGGCCATGGCA 2801

Db     1528 GTGGATGTCTGCAAGGCTGGTGATCGGGTCGAAGTGACCGGTATTTTCCGGTGCAACCCT 1587

Qy     2802 GTTCGAGTTGGTCCGAATCAACGAACATTACGAGCATTGTATAAGACCTACATCGATTGC 2861

Db     1588 GTGCGCGTTAATCCTCGCCAGCGTACACAGAAGTCGCTGTTCAAGACGTACATAGATGTT 1647

Qy     2862 GTGCACGTCAAGAAGTCTGACAGGGGTCGACTGCAAACTGAAGATCCTATGGAGATGGAT 2921

Db     1648 CTTCATGTTCAGAAGATCGATCGCAAGAAGTTGGGTATCGACGTCTCGACCATCGAGCAG 1707

Qy     2922 AAGGAGAATGATATGTATGCTGGGTATCATGAAAGTGATACTTCAGAAGCTGCTAATGAA 2981

Db     1708 GAGCTCTCGGAACAGGCGGCTGGGGATGCAGAACAAACACGTAGGCTCACTGCGGAGGAG 1767

Qy     2982 GCAAAGATTCAAAAACTTAAAGAGCTGTCCAAGCTCCCGGGCATTTATGATAGACTTTCA 3041

Db     1768 GAAGAG------AAAATTAAGCGAACTGCTACCAGACCTGATCTGTATGAGCTTCTCTCT 1821

Qy     3042 AGGTCGCTGGCTCCAAGCATTTGGGAGCTTGAAGATATTAAAAAGGGTCTTCTTTGCCAG 3101

Db     1822 CGGTCCTTGGCCCCCAGCATCTACGAGATGGACGACGTGAAGAAGGGAATCCTGCTTCAG 1881

Qy     3102 CTCTTTGGTGGGAAGGCTAAGAAAATTCCATCTGGAG------CATCTTTCCGAGGTGAC 3155

Db     1882 TTGTTTGGAGGCACCAACAAGACCTTCCAGAAGGGTGGTAACCCACGATACCGTGGAGAT 1941

Qy     3156 ATCAATGTTTTACTTGTTGGGGACCCTGGTACCAGTAAATCTCAGCTGCTTCAGTATGTG 3215

Db     1942 ATCAATATCCTTCTCTGTGGTGACCCATCTACATCCAAGTCCCAGCTTCTTCGTTACGTC 2001

Qy     3216 CACAAGATAGCTCCTCGTGGAATCTACACTAGTGGGCGAGGAAGTTCGGCGGTTGGGCTG 3275

Db     2002 CATAAGATTGCCCCTCGCGGTGTGTATACCAGCGGCAAGGGCTCCTCGGCTGTTGGTCTT 2061

Qy     3276 ACAGCGTATGTAACGAAGGATCCAGAAACTCGAGAGACGGTATTGGAGAGCGGAGCTTTG 3335

Db     2062 ACGGCGTACGTCACCCGCGATCCTGAAACCCGCCAGATGGTCCTCGAGTCGGGTGCCTTG 2121

Qy     3336 GTTCTTAGTGATCGTGGGATATGCTGTATCGATGAGTTCGACAAAATGTCTGATAATGCC 3395
```

```
Db     2122 GTTCTTTCAGACGGCGGTATCTGTTGCATCGACGAGTTCGACAAGATGAACGAATCCACT 2181

Qy     3396 CGAAGCATGCTTCATGAGGTAATGGAGCAACAAACGGTATCTGTAGCCAAAGGGGGTATC 3455
Db     2182 CGGTCCGTTCTGCATGAAGTCATGGAACAACAGACAGTATCTATCGCCAAGGCAGGCATT 2241

Qy     3456 ATTGCCTCGCTGAACGCTCGGACGTCTGTCCTTGCATGTGCAAATCCTAGTGGGTCCCGA 3515
Db     2242 ATCACTACTTTGAACGCTAGGACCAGCATCCTGGCTTCCGCCAATCCGATCGGTAGCAGG 2301

Qy     3516 TACAATGCGCGCCTTTCTGTGATTGATAACATCCAGCTTCCTCCAACTCTACTTTCTAGA 3575
Db     2302 TACAATCCCAACTTGCCCGTTCCTCAAAATATTGACCTTCCGCCTACCTTGCTCTCCCGA 2361

Qy     3576 TTTGATTTAATTTACTTAATGCTCGACAAACCAGACGAGCAAAACGATCGTCGTCTCGCC 3635
Db     2362 TTCGACTTGGTATACCTCGTGCTGGACCGAGTGGATGAGCAGGAAGATCGTCGGCTCGCT 2421

Qy     3636 AGGCATCTCGTGGCTTTACACTATGAAAAC------TATGAAGTTTCAAAGCAGGACGCC 3689
Db     2422 AAGCATCTTGTCAATATGTACCTGGAAGACAGACCTGAGCATGCTGCCGAGCAAGAAATC 2481

Qy     3690 TTAGATCTACAAACACTTACCGCGTATATCACCTATGCTCGTCAGCATGTACATCCTACA 3749
Db     2482 TTGCCGATCGAATTCCTTACAGCCTATATCACCTACGCCAAGACCAAAGTCCATCCAGTG 2541

Qy     3750 TTAAGTGATGAAGCTGCTGAAGATTTGATTAATGGCTATGTTGAGATGCGCCAAAAGGGC 3809
Db     2542 CTCACACCGGCCGCCGGTAAAGCCTTGTCGGATGCTTACGTTAACATGCGTAAGCTTGGA 2601

Qy     3810 AACTTTCCTGGAAGCAGTAAAAAGGTGATAACAGCCACACCTCGGCAACTCGAAAGTATG 3869
Db     2602 GATGACATCCGGTCTTCTGACCGCCGTATCACCGCTACCACTCGTCAACTGGAGTCCATG 2661

Qy     3870 ATTCGTATCAGTGAAGCCCTAGCTCGAATGAGATTTTCTGAAGTGGTAGAGAAAGTTGAT 3929
Db     2662 ATCCGACTGTCGGAAGCGCATGCGCGTATGCGGCTATCGCCGGAGGTCACTGCGGATGAT 2721

Qy     3930 GCAGCAGAAGCTGTGCGCCTTTTAGACGTCGCTTTGCAGCAATCTGCTACTGATCATGCA 3989
Db     2722 GTGGAGGAAGCCGTGCGCCTGATCCGCTCCGCCATCAAGCAGGCGGCCACTGACTCTCGG 2781

Qy     3990 ACAGGTACGATAGACATGGATCTTATCACGACTGGAGTGTCGGCCAGCGAGCGTATTCGT 4049
Db     2782 ACCGGTCTGATCGACATGAGCTTGTTGACGGAGGGCACTAGTGCCAGCGAGAGACGCAGC 2841

Qy     4050 CGGG 4053
Db     2842 CGGG 2845
```

Search completed: December 6, 2005, 09:52:34
Job time : 2140 secs

```
OM nucleic - nucleic search, using sw model

Run on:          December  6, 2005, 02:17:11 ; Search time 395 Seconds
                                              (without alignments)
                                              3425.784 Million cell updates/sec

Title:           US-10-768-511-5
Perfect score:   4348
Sequence:        1 atggcgcgccgcactcacgt..........gcgcaattcccagagctcgc 4348

Scoring table:   IDENTITY_NUC
                 Gapop 10.0 , Gapext 1.0

Searched:        3289935 seqs, 155610033 residues

Total number of hits satisfying chosen parameters:      6579870

Minimum DB seq length: 0
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%
                 Maximum Match 100%
                 Listing first 45 summaries

Database :       Published_Applications_NA_New:*
                 1:  /cgn2_6/ptodata/1/pubpna/US09_NEW_PUB.seq:*
                 2:  /cgn2_6/ptodata/1/pubpna/US06_NEW_PUB.seq:*
                 3:  /cgn2_6/ptodata/1/pubpna/US07_NEW_PUB.seq:*
                 4:  /cgn2_6/ptodata/1/pubpna/US08_NEW_PUB.seq:*
                 5:  /cgn2_6/ptodata/1/pubpna/PCT_NEW_PUB.seq:*
                 6:  /cgn2_6/ptodata/1/pubpna/US10_NEW_PUB.seq:*
                 7:  /cgn2_6/ptodata/1/pubpna/US11_NEW_PUB.seq:*
                 8:  /cgn2_6/ptodata/1/pubpna/US11_NEW_PUB.seq2:*
                 9:  /cgn2_6/ptodata/1/pubpna/US11_NEW_PUB.seq3:*
                 10:  /cgn2_6/ptodata/1/pubpna/US60_NEW_PUB.seq:*
```

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

	Result No.	Score	% Query Match	Length	DB	ID	Description
	1	186.4	4.3	3769	7	US-11-087-227-89	Sequence 89, Appl
	2	180	4.1	2821	7	US-11-087-227-13	Sequence 13, Appl
	3	180	4.1	2900	7	US-11-087-227-15	Sequence 15, Appl
	4	173.2	4.0	2715	6	US-10-507-275-4	Sequence 4, Appli
	5	160.8	3.7	3453	7	US-11-087-227-11	Sequence 11, Appl
	6	157.6	3.6	3379	6	US-10-507-275-2	Sequence 2, Appli
	7	123.2	2.8	507	7	US-11-108-172-158	Sequence 158, App
	8	58.2	1.3	1851	6	US-10-750-185-45476	Sequence 45476, A
	9	54.8	1.3	1312	6	US-10-750-185-55669	Sequence 55669, A
	10	39.6	0.9	876	6	US-10-750-185-27656	Sequence 27656, A
	11	37.6	0.9	449	6	US-10-821-234-393	Sequence 393, App
c	12	35.6	0.8	978	6	US-10-821-234-409	Sequence 409, App
c	13	35.4	0.8	3150	6	US-10-793-626-4034	Sequence 4034, Ap
	14	35.4	0.8	4231	6	US-10-793-626-4115	Sequence 4115, Ap
	15	34.2	0.8	1263	6	US-10-793-626-3317	Sequence 3317, Ap
c	16	34.2	0.8	1410	6	US-10-750-185-62786	Sequence 62786, A
	17	34.2	0.8	2247	6	US-10-750-185-48673	Sequence 48673, A
c	18	34.2	0.8	4116	6	US-10-793-626-4100	Sequence 4100, Ap
c	19	34	0.8	1210	6	US-10-750-185-41443	Sequence 41443, A
	20	33.4	0.8	2957	6	US-10-793-626-4135	Sequence 4135, Ap
c	21	33	0.8	1758	6	US-10-750-185-38232	Sequence 38232, A
c	22	33	0.8	2428	6	US-10-750-185-50035	Sequence 50035, A
c	23	32.8	0.8	1003	6	US-10-750-185-37272	Sequence 37272, A
	24	32.4	0.7	1359	6	US-10-793-626-3091	Sequence 3091, Ap
	25	32.4	0.7	3053	6	US-10-793-626-3514	Sequence 3514, Ap
c	26	32.4	0.7	3253	6	US-10-793-626-3887	Sequence 3887, Ap
	27	32	0.7	1522	6	US-10-750-185-35384	Sequence 35384, A
c	28	32	0.7	2463	6	US-10-750-185-55745	Sequence 55745, A
c	29	31.8	0.7	1907	6	US-10-750-185-41934	Sequence 41934, A
c	30	31.8	0.7	4320	6	US-10-750-185-27669	Sequence 27669, A
c	31	31.6	0.7	1298	6	US-10-750-185-51133	Sequence 51133, A
c	32	31.4	0.7	600	6	US-10-750-185-20304	Sequence 20304, A
	33	31.4	0.7	1106	6	US-10-750-185-37114	Sequence 37114, A
c	34	31.4	0.7	1144	6	US-10-750-185-46616	Sequence 46616, A
	35	31.4	0.7	1773	6	US-10-750-185-42030	Sequence 42030, A
c	36	31.4	0.7	1894	6	US-10-750-185-64107	Sequence 64107, A
c	37	31.4	0.7	3238	6	US-10-750-185-38644	Sequence 38644, A
	38	31.4	0.7	3666	6	US-10-793-626-3543	Sequence 3543, Ap
c	39	31.4	0.7	5468	6	US-10-821-234-49	Sequence 49, Appl
	40	31.4	0.7	5571	6	US-10-750-185-53588	Sequence 53588, A
c	41	31.4	0.7	171162	7	US-11-112-908-38	Sequence 38, Appl
c	42	31.2	0.7	1987	6	US-10-750-185-49778	Sequence 49778, A
c	43	31.2	0.7	3552	6	US-10-750-185-42179	Sequence 42179, A
c	44	31	0.7	944	6	US-10-750-185-25432	Sequence 25432, A
c	45	31	0.7	1276	6	US-10-750-185-37419	Sequence 37419, A

ALIGNMENTS

```
RESULT 1
US-11-087-227-89
; Sequence 89, Application US/11087227
; Publication No. US20050260566A1
; GENERAL INFORMATION:
;  APPLICANT: Fischer, Timothy J.
;  APPLICANT:  Malinowski, Douglas P.
;  APPLICANT:  Taylor, Adriann J.
;  APPLICANT:  Parker, Margaret R.
;  TITLE OF INVENTION: METHODS AND COMPOSITIONS FOR THE
;  TITLE OF INVENTION:  DETECTION OF CERVICAL DISEASE
;  FILE REFERENCE: 046143/287139
;  CURRENT APPLICATION NUMBER: US/11/087,227
;  CURRENT FILING DATE:  2005-03-23
;  PRIOR APPLICATION NUMBER: 60/556,495
;  PRIOR FILING DATE: 2004-03-24
;  NUMBER OF SEQ ID NOS: 90
;  SOFTWARE: FastSEQ for Windows Version 4.0
; SEQ ID NO 89
;   LENGTH: 3769
;   TYPE: DNA
;   ORGANISM: Homo sapiens
US-11-087-227-89

  Query Match             4.3%;  Score 186.4;  DB 7;  Length 3769;
  Best Local Similarity   51.7%;  Pred. No. 3.2e-50;
  Matches  477;  Conservative    0;  Mismatches  436;  Indels    9;  Gaps    2;

Qy       3025 TTTATGATAGACTTTCAAGGTCGCTGGCTCCAAGCATTTGGGAGCTTGAAGATATTAAAA 3084

Db       1113 TATACCACAATCTTTGTACCAGCCTGTTCCCTACTATACATGGCAATGATGAAGTAAAAC 1172

Qy       3085 AGGGTCTTCTTTGCCAGCTCTTTGGTGGGAAGGCTAAGAAAATTCCATCTGGAGCATCTT 3144

Db       1173 GGGGTGTCCTGCTGATGCTCTTTGGTGGCGTTCCAAAGACAACAGGAGAAGGACCTCTC 1232

Qy       3145 TCCGAGGTGACATCAATGTTTTACTTGTTGGGGACCCTGGTACCAGTAAATCTCAGCTGC 3204

Db       1233 TTCGAGGGGACATAAATGTTTGCATTGTTGGTGACCCAAGTACAGCTAAGAGCCAATTTC 1292

Qy       3205 TTCAGTATGTGCACAAGATAGCTCCTCGTGGAATCTACACTAGTGGGCGAGGAAGTTCGG 3264

Db       1293 TCAAGCACGTGGAGGAGTTCAGCCCCAGAGCTGTCTACACCAGTGGTAAAGCGTCCAGTG 1352

Qy       3265 CGGTTGGGCTGACAGCGTATGTAACGAAGGATCCAGAAACTCGAGAGACGGTATTGGAGA 3324
```

```
Db   1353 CTGCTGGCTTAACAGCAGCTGTTGTGAGAGATGAAGAATCTCATGAGTTTGTCATTGAGG 1412

Qy   3325 GCGGAGCTTTGGTTCTTAGTGATCGTGGGATATGCTGTATCGATGAGTTCGACAAAATGT 3384
Db   1413 CTGGAGCTTTGATGTTGGCTGATAATGGTGTGTGTTGTATTGATGAATTTGATAAGATGG 1472

Qy   3385 CTGATAATGCCCGAAGCATGCTTCATGAGGTAATGGAGCAACAAACGGTATCTGTAGCCA 3444
Db   1473 ACGTGCGGGATCAAGTTGCTATTCATGAAGCTATGGAACAGCAGACCATATCCATCACTA 1532

Qy   3445 AAGGGGGTATCATTGCCTCGCTGAACGCTCGGACGTCTGTCCTTGCATGTGCAAATCCTA 3504
Db   1533 AAGCAGGAGTGAAGGCTACTCTGAACGCCCGGACGTCCATTTTGGCAGCAGCAAACCCAA 1592

Qy   3505 GTGGGTCCCGATACAATGCGCGCCTTTCTGTGATTGATAACATCCAGCTTCCTCCAACTC 3564
Db   1593 TCAGTGGACACTATGACAGATCAAAATCATTGAAACAGAATATAAATTTGTCAGCTCCCA 1652

Qy   3565 TACTTTCTAGATTTGATTTAATTTACTTAATGCTCGACAAACCAGACGAGCAAAACGATC 3624
Db   1653 TCATGTCCCGATTCGATCTCTTCTTTATCCTTGTGGATGAATGTAATGAGGTTACAGATT 1712

Qy   3625 GTCGTCTCGCCAGGCATCTCGTGGCTTTACACTATGAAAACTATGAAGTTTCAAAGCAGG 3684
Db   1713 ATGCCATTGCCAGGCGCATAGTAGATTTGC---ATTCAAGAATTGAGGAATCAATTGATC 1769

Qy   3685 ACGCCTTAGATCTACAAACACTTACCGCGTATATCACCTATGCTCGTCAGCATGTACATC 3744
Db   1770 GTGTCTATTCCCTCGATGATATCAGAAGATATCTTCTCTTTGCAAGACAGTTTAAAC--- 1826

Qy   3745 CTACATTAAGTGATGAAGCTGCTGAAGATTTGATTAATGGCTATGTTGAGATGCGCCAAA 3804
Db   1827 ---CCAAGATTTCCAAAGAGTCAGAGGACTTCATTGTGGAGCAATATAAACATCTCCGCC 1883

Qy   3805 AGGGCAACTTTCCTGGAAGCAGTAAAAAGGTGATAACAGCCACACCTCGGCAACTCGAAA 3864
Db   1884 AGAGAGATGGTTCTGGAGTGACCAAGTCTTCATGGAGGATTACAGTGCGACAGCTTGAGA 1943

Qy   3865 GTATGATTCGTATCAGTGAAGCCCTAGCTCGAATGAGATTTTCTGAAGTGGTAGAGAAAG 3924
Db   1944 GCATGATTCGTCTCTCTGAAGCTATGGCTCGGATGCACTGCTGTGATGAGGTCCAACCTA 2003

Qy   3925 TTGATGCAGCAGAAGCTGTGCG 3946
Db   2004 AACATGTGAAGGAAGCTTTCCG 2025
```

RESULT 2
US-11-087-227-13
; Sequence 13, Application US/11087227
; Publication No. US20050260566A1
; GENERAL INFORMATION:
; APPLICANT: Fischer, Timothy J.
; APPLICANT: Malinowski, Douglas P.
; APPLICANT: Taylor, Adriann J.
; APPLICANT: Parker, Margaret R.
; TITLE OF INVENTION: METHODS AND COMPOSITIONS FOR THE
; TITLE OF INVENTION: DETECTION OF CERVICAL DISEASE
; FILE REFERENCE: 046143/287139
; CURRENT APPLICATION NUMBER: US/11/087,227
; CURRENT FILING DATE: 2005-03-23
; PRIOR APPLICATION NUMBER: 60/556,495
; PRIOR FILING DATE: 2004-03-24
; NUMBER OF SEQ ID NOS: 90
; SOFTWARE: FastSEQ for Windows Version 4.0
; SEQ ID NO 13
; LENGTH: 2821
; TYPE: DNA
; ORGANISM: Homo sapiens
US-11-087-227-13

Query Match 4.1%; Score 180; DB 7; Length 2821;
Best Local Similarity 51.2%; Pred. No. 3.3e-48;
Matches 479; Conservative 0; Mismatches 445; Indels 12; Gaps 2;

```
Qy   3020 GGGCATTTATGATAGACTTTCAAGGTCGCTGGCTCCAAGCATTTGGGAGCTTGAAGATAT 3079
Db   1500 GGATTTCTACGAAAAGCTGGCAGCTTCAATCGCCCCAGAAATATACGGGCATGAAGATGT 1559

Qy   3080 TAAAAAGGGTCTTCTTTGCCAGCTCTTTGGTGGGAAGGCTAAGAAAATTCCATCTGGAGC 3139
Db   1560 GAAGAAGGCACTGCTGCTCCTGCTAGTCGGGGGTGTGG---ACCAGTCTCCTCGAGGCAT 1616

Qy   3140 ATCTTTCCGAGGTGACATCAATGTTTTACTTGTTGGGGACCCTGGTACCAGTAAATCTCA 3199
Db   1617 GAAAATCCGGGGCAACATCAACATCTGTCTGATGGGGGATCCTGGTGTGGCCAAGTCTCA 1676

Qy   3200 GCTGCTTCAGTATGTGCACAAGATAGCTCCTCGTGGAATCTACACTAGTGGGCGAGGAAG 3259
Db   1677 GCTCCTGTCATACATTGATCGACTGGCGCCTCGCAGCCAGTACACAACAGGCCGGGGCTC 1736

Qy   3260 TTCGGCGGTTGGGCTGACAGCGTATGTAACGAAGGATCCAGAAACTCGAGAGACGGTATT 3319
Db   1737 CTCAGGAGTGGGGCTTACGGCAGCTGTGCTGAGAGACTCCGTGAGTGGAGAACTGACCTT 1796

Qy   3320 GGAGAGCGGAGCTTTGGTTCTTAGTGATCGTGGGATATGCTGTATCGATGAGTTCGACAA 3379
Db   1797 AGAGGGTGGGGCCCTGGTGCTGGCTGACCAGGGTGTGTGCTGCATTGATGAGTTCGACAA 1856

Qy   3380 AATGTCTGATAATGCCCGAAGCATGCTTCATGAGGTAATGGAGCAACAAACGGTATCTGT 3439
Db   1857 GATGGCTGAGGCCGACCGCACAGCCATCCACGAGGTCATGGAGCAGCAGACCATCTCCAT 1916

Qy   3440 AGCCAAAGGGGGTATCATTGCCTCGCTGAACGCTCGGACGTCTGTCCTTGCATGTGCAAA 3499
Db   1917 TGCCAAGGCCGGCATTCTCACCACACTCAATGCCCGCTGCTCCATCCTGGCTGCCGCCAA 1976

Qy   3500 TCCTAGTGGGTCCCGATACAATGCGCGCCTTTCTGTGATTGATAACATCCAGCTTCCTCC 3559
Db   1977 CCCTGCCTACGGGCGCTACAACCCTCGCCGCAGCCTGGAGCAGAACATACAGCTACCTGC 2036

Qy   3560 AACTCTACTTTCTAGATTTGATTTAATTTACTTAATGCTCGACAAACCAGACGAGCAAAA 3619
Db   2037 TGCACTGCTCTCCCGGTTTGACCTCCTCTGGCTGATTCAGGACCGGCCCGACCGAGACAA 2096

Qy   3620 CGATCGTCGTCTCGCCAGGCATCTCGTGGCTTTACACTATGAAAACTATGAAGTTTCAAA 3679
Db   2097 TGACCTACGGTTGGCCCAGCACATCACCTATGTGCACCAGCACAGCCGGCAGCCCCCCTC 2156

Qy   3680 GCAGGACGCCTTAGATCTACAAACACTTACCGCGTATATCACCTATGCTCGTCAGCATGT 3739
Db   2157 CCAGTTTGAACCTCTGGACATGAAGCTCATGAGGCGTTACATAGCCATGTGCCGCGAGAA 2216

Qy   3740 ACATCCTACATTAAGTGATGAAGCTGCTGAAGATTTGATTAATGGCTATGTTGAGATGCG 3799
Db   2217 GCAGCCCATGGTGCCAGAGTCTCTGGCTGACTACATCACAGCAGCATACGTGGAGATGAG 2276

Qy   3800 CCAAAAGGGCAACTTTCCTGGAAGCAGTAAAAAGGTGATAACAGCCACACCTCGGCAACT 3859
Db   2277 GCGAGAGG---------CTTGGGCTAGTAAGGATGCCACCTATACTTCTGCCCGGACCCT 2327

Qy   3860 CGAAAGTATGATTCGTATCAGTGAAGCCCTAGCTCGAATGAGATTTTCTGAAGTGGTAGA 3919
Db   2328 GCTGGCTATCCTGCGCCTTTCCACTGCTCTGGCACGTCTGAGAATGGTGGATGTGGTGGA 2387

Qy   3920 GAAAGTTGATGCAGCAGAAGCTGTGCGCCTTTTAGA 3955
Db   2388 GAAAGAAGATGTGAATGAAGCCATCAGGCTAATGGA 2423
```

RESULT 3
US-11-087-227-15
; Sequence 15, Application US/11087227
; Publication No. US20050260566A1
; GENERAL INFORMATION:
; APPLICANT: Fischer, Timothy J.

```
; APPLICANT:  Malinowski, Douglas P.
; APPLICANT:  Taylor, Adriann J.
; APPLICANT:  Parker, Margaret R.
; TITLE OF INVENTION: METHODS AND COMPOSITIONS FOR THE
; TITLE OF INVENTION:  DETECTION OF CERVICAL DISEASE
; FILE REFERENCE: 046143/287139
; CURRENT APPLICATION NUMBER: US/11/087,227
; CURRENT FILING DATE:  2005-03-23
; PRIOR APPLICATION NUMBER: 60/556,495
; PRIOR FILING DATE: 2004-03-24
; NUMBER OF SEQ ID NOS: 90
; SOFTWARE: FastSEQ for Windows Version 4.0
; SEQ ID NO 15
;   LENGTH: 2900
;   TYPE: DNA
;   ORGANISM: Homo sapiens
US-11-087-227-15

  Query Match             4.1%;  Score 180;  DB 7;  Length 2900;
  Best Local Similarity   51.2%;  Pred. No. 3.3e-48;
  Matches  479;  Conservative    0;  Mismatches  445;  Indels   12;  Gaps    2;

Qy       3020 GGGCATTTATGATAGACTTTCAAGGTCGCTGGCTCCAAGCATTTGGGAGCTTGAAGATAT 3079
Db       1579 GGATTTCTACGAAAAGCTGGCAGCTTCAATCGCCCCAGAAATATACGGGCATGAAGATGT 1638

Qy       3080 TAAAAAGGGTCTTCTTTGCCAGCTCTTTGGTGGGAAGGCTAAGAAAATTCCATCTGGAGC 3139
Db       1639 GAAGAAGGCACTGCTGCTCCTGCTAGTCGGGGGTGTGG---ACCAGTCTCCTCGAGGCAT 1695

Qy       3140 ATCTTTCCGAGGTGACATCAATGTTTTACTTGTTGGGGACCCTGGTACCAGTAAATCTCA 3199
Db       1696 GAAAATCCGGGGCAACATCAACATCTGTCTGATGGGGGATCCTGGTGTGGCCAAGTCTCA 1755

Qy       3200 GCTGCTTCAGTATGTGCACAAGATAGCTCCTCGTGGAATCTACACTAGTGGGCGAGGAAG 3259
Db       1756 GCTCCTGTCATACATTGATCGACTGGCGCCTCGCAGCCAGTACACAACAGGCCGGGGCTC 1815

Qy       3260 TTCGGCGGTTGGGCTGACAGCGTATGTAACGAAGGATCCAGAAACTCGAGAGACGGTATT 3319
Db       1816 CTCAGGAGTGGGGCTTACGGCAGCTGTGCTGAGAGACTCCGTGAGTGGAGAACTGACCTT 1875

Qy       3320 GGAGAGCGGAGCTTTGGTTCTTAGTGATCGTGGGATATGCTGTATCGATGAGTTCGACAA 3379
Db       1876 AGAGGGTGGGGCCCTGGTGCTGGCTGACCAGGGTGTGCTGCATTGATGAGTTCGACAA 1935

Qy       3380 AATGTCTGATAATGCCCGAAGCATGCTTCATGAGGTAATGGAGCAACAAACGGTATCTGT 3439
Db       1936 GATGGCTGAGGCCGACCGCACAGCCATCCACGAGGTCATGGAGCAGCAGACCATCTCCAT 1995

Qy       3440 AGCCAAAGGGGGTATCATTGCCTCGCTGAACGCTCGGACGTCTGTCCTTGCATGTGCAAA 3499
Db       1996 TGCCAAGGCCGGCATTCTCACCACACTCAATGCCCGCTGCTCCATCCTGGCTGCCGCCAA 2055

Qy       3500 TCCTAGTGGGTCCCGATACAATGCGCGCCTTTCTGTGATTGATAACATCCAGCTTCCTCC 3559
Db       2056 CCCTGCCTACGGGCGCTACAACCCTCGCCGCAGCCTGGAGCAGAACATACAGCTACCTGC 2115

Qy       3560 AACTCTACTTTCTAGATTTGATTTAATTTACTTAATGCTCGACAAACCAGACGAGCAAAA 3619
Db       2116 TGCACTGCTCTCCCGGTTTGACCTCCTCTGGCTGATTCAGGACCGGCCCGACCGAGACAA 2175

Qy       3620 CGATCGTCGTCTCGCCAGGCATCTCGTGGCTTTACACTATGAAAACTATGAAGTTTCAAA 3679
Db       2176 TGACCTACGGTTGGCCCAGCACATCACCTATGTGCACCAGCACAGCCGGCAGCCCCCCTC 2235

Qy       3680 GCAGGACGCCTTAGATCTACAAACACTTACCGCGTATATCACCTATGCTCGTCAGCATGT 3739
Db       2236 CCAGTTTGAACCTCTGGACATGAAGCTCATGAGGCGTTACATAGCCATGTGCCGCGAGAA 2295

Qy       3740 ACATCCTACATTAAGTGATGAAGCTGCTGAAGATTTGATTAATGGCTATGTTGAGATGCG 3799
Db       2296 GCAGCCCATGGTGCCAGAGTCTCTGGCTGACTACATCACAGCAGCATACGTGGAGATGAG 2355

Qy       3800 CCAAAAGGGCAACTTTCCTGGAAGCAGTAAAAAGGTGATAACAGCCACACCTCGGCAACT 3859
Db       2356 GCGAGAGG---------CTTGGGCTAGTAAGGATGCCACCTATACTTCTGCCCGGACCCT 2406

Qy       3860 CGAAAGTATGATTCGTATCAGTGAAGCCCTAGCTCGAATGAGATTTTCTGAAGTGGTAGA 3919
Db       2407 GCTGGCTATCCTGCGCCTTTCCACTGCTCTGGCACGTCTGAGAATGGTGGATGTGGTGGA 2466

Qy       3920 GAAAGTTGATGCAGCAGAAGCTGTGCGCCTTTTAGA 3955
Db       2467 GAAAGAAGATGTGAATGAAGCCATCAGGCTAATGGA 2502


RESULT 4
US-10-507-275-4
; Sequence 4, Application US/10507275
; Publication No. US20050250166A1
; GENERAL INFORMATION:
;  APPLICANT: Masai, Hisao
;  APPLICANT:  Tamai, Katsuyuki
;  APPLICANT:  Medical and Biological Laboratories Co., Ltd.
;  APPLICANT:  Japan Science and Technology Agency
;  APPLICANT:  Ginkgo Biomedical Research Institute Co., Ltd.
;  TITLE OF INVENTION: Cdc7-ASK Kinase Complex, Substrates of the Kinase Complex,
;  TITLE OF INVENTION:  Specific Antibodies to the Substrates, and Screening Methods
;  TITLE OF INVENTION:  Using the Same to Screen for Compounds Comprising Cdc7-ASK
;  TITLE OF INVENTION:  Kinase Inhititory Ability
;  FILE REFERENCE: 082368-001100US
;  CURRENT APPLICATION NUMBER: US/10/507,275
;  CURRENT FILING DATE:  2004-09-09
;  PRIOR APPLICATION NUMBER: JP 2002-067702
;  PRIOR FILING DATE: 2002-03-12
;  PRIOR APPLICATION NUMBER: WO PCT/JP03/02918
;  PRIOR FILING DATE: 2003-03-12
;  NUMBER OF SEQ ID NOS: 21
;  SOFTWARE: PatentIn Ver. 2.1
; SEQ ID NO 4
;   LENGTH: 2715
;   TYPE: DNA
;   ORGANISM: Mus musculus
;   FEATURE:
;   NAME/KEY: CDS
;   LOCATION: (1)..(2715)
US-10-507-275-4

  Query Match             4.0%;  Score 173.2;  DB 6;  Length 2715;
  Best Local Similarity   53.4%;  Pred. No. 5.5e-46;
  Matches  364;  Conservative    0;  Mismatches  318;  Indels    0;  Gaps    0;

Qy       3006 CTGTCCAAGCTCCCGGGCATTTATGATAGACTTTCAAGGTCGCTGGCTCCAAGCATTTGG 3065
Db       1399 CTCTCCAAGGATCAGCAAATTGGAGAGAAGATCTTTGCCAGCATTGCACCCTCCATCTAT 1458

Qy       3066 GAGCTTGAAGATATTAAAAAGGGTCTTCTTTGCCAGCTCTTTGGTGGGAAGGCTAAGAAA 3125
Db       1459 GGGCATGAAGACATCAAGAGAGGCCTGGCTCTGGCCCTGTTTGGAGGGGAGCCCAAGAAC 1518

Qy       3126 ATTCCATCTGGAGCATCTTTCCGAGGTGACATCAATGTTTTACTTGTTGGGGACCCTGGT 3185
Db       1519 CCAGGTGGAAAGCACAAGGTTCGAGGTGACATTAATGTGCTCTTGTGTGGGGACCCTGGC 1578

Qy       3186 ACCAGTAAATCTCAGCTGCTTCAGTATGTGCACAAGATAGCTCCTCGTGGAATCTACACT 3245
Db       1579 ACAGCAAAGTCCCAATTCCTCAAATACATCGAGAAAGTGTCTAGCCGTGCCATCTTCACC 1638

Qy       3246 AGTGGGCGAGGAAGTTCGGCGGTTGGGCTGACAGCGTATGTAACGAAGGATCCAGAAACT 3305
Db       1639 ACTGGCCAGGGTGCGTCAGCAGTGGGTCTCACCGCGTACGTTCAGCGGCATCCCGTCAGC 1698

Qy       3306 CGAGAGACGGTATTGGAGAGCGGAGCTTTGGTTCTTAGTGATCGTGGGATATGCTGTATC 3365
Db       1699 AGAGAGTGGACCTTAGAGGCGGGAGCCCTGGTTCTGGCTGACCGGGGGGTGTGTCTCATT 1758
```

```
Qy       3366 GATGAGTTCGACAAAATGTCTGATAATGCCCGAAGCATGCTTCATGAGGTAATGGAGCAA 3425
Db       1759 GACGAGTTTGACAAGATGAATGACCAGGACAGGACCAGCATCCACGAGGCCATGGAACAG 1818

Qy       3426 CAAACGGTATCTGTAGCCAAAGGGGGTATCATTGCCTCGCTGAACGCTCGGACGTCTGTC 3485
Db       1819 CAAAGCATCTCCATCTCCAAGGCTGGCATCGTTACCTCGCTGCAAGCCCGCTGCACTGTC 1878

Qy       3486 CTTGCATGTGCAAATCCTAGTGGGTCCCGATACAATGCGCGCCTTTCTGTGATTGATAAC 3545
Db       1879 ATAGCTGCTGCCAACCCCATAGGAGGCCGCTACGACCCTTCACTGACCTTCTCAGAGAAT 1938

Qy       3546 ATCCAGCTTCCTCCAACTCTACTTTCTAGATTTGATTTAATTTACTTAATGCTCGACAAA 3605
Db       1939 GTAGACCTCACAGAGCCCATCATTTCCCGCTTTGATGTCCTGTGTGTGGTGAGGGACACT 1998

Qy       3606 CCAGACGAGCAAAACGATCGTCGTCTCGCCAGGCATCTCGTGGCTTTACACTATGAAAAC 3665
Db       1999 GTTGATCCAGTTCAGGATGAGATGCTGGCCCGCTTTGTGGTTGGCAGCCACGTCAGACAC 2058

Qy       3666 TATGAAGTTTCAAAGCAGGACG 3687
Db       2059 CACCCCAGTAACAAGAAGGATG 2080
```

RESULT 5
US-11-087-227-11
; Sequence 11, Application US/11087227
; Publication No. US20050260566A1
; GENERAL INFORMATION:
; APPLICANT: Fischer, Timothy J.
; APPLICANT: Malinowski, Douglas P.
; APPLICANT: Taylor, Adriann J.
; APPLICANT: Parker, Margaret R.
; TITLE OF INVENTION: METHODS AND COMPOSITIONS FOR THE
; TITLE OF INVENTION: DETECTION OF CERVICAL DISEASE
; FILE REFERENCE: 046143/287139
; CURRENT APPLICATION NUMBER: US/11/087,227
; CURRENT FILING DATE: 2005-03-23
; PRIOR APPLICATION NUMBER: 60/556,495
; PRIOR FILING DATE: 2004-03-24
; NUMBER OF SEQ ID NOS: 90
; SOFTWARE: FastSEQ for Windows Version 4.0
; SEQ ID NO 11
;   LENGTH: 3453
;   TYPE: DNA
;   ORGANISM: Homo sapiens
US-11-087-227-11

```
  Query Match             3.7%;  Score 160.8;  DB 7;  Length 3453;
  Best Local Similarity   50.9%;  Pred. No. 7.8e-42;
  Matches  381;  Conservative    0;  Mismatches  367;  Indels    0;  Gaps    0;

Qy       2940 GCTGGGTATCATGAAAGTGATACTTCAGAAGCTGCTAATGAAGCAAAGATTCAAAAACTT 2999
Db       1390 GCCAAGAAGGACAACAAGGTTGCTGTAGGGGAACTGACCGATGAAGATGTGAAGATGATC 1449

Qy       3000 AAAGAGCTGTCCAAGCTCCCGGGCATTTATGATAGACTTTCAAGGTCGCTGGCTCCAAGC 3059
Db       1450 ACTAGCCTCTCCAAGGATCAGCAGATCGGAGAGAAGATCTTTGCCAGCATTGCTCCTTCC 1509

Qy       3060 ATTTGGAGCTTGAAGATATTAAAAAGGGTCTTCTTTGCCAGCTCTTTGGTGGGAAGGCT 3119
Db       1510 ATCTATGGTCATGAAGACATCAAGAGAGGCCTGGCTCTGGCCCTGTTCGGAGGGGAGCCC 1569

Qy       3120 AAGAAAATTCCATCTGGAGCATCTTTCCGAGGTGACATCAATGTTTTACTTGTTGGGGAC 3179
Db       1570 AAAAACCCAGGTGGCAAGCACAAGGTACGTGGTGATATCAACGTGCTCTTGTGCGGAGAC 1629

Qy       3180 CCTGGTACCAGTAAATCTCAGCTGCTTCAGTATGTGCACAAGATAGCTCCTCGTGGAATC 3239
Db       1630 CCTGGCACAGCGAAGTCGCAGTTTCTCAAGTATATTGAGAAAGTGTCCAGCCGAGCCATC 1689

Qy       3240 TACACTAGTGGGCGAGGAAGTTCGGCGGTTGGGCTGACAGCGTATGTAACGAAGGATCCA 3299
Db       1690 TTCACCACTGGCCAGGGGGCGTCGGCTGTGGGCCTCACGGCGTATGTCCAGCGGCACCCT 1749

Qy       3300 GAAACTCGAGAGACGGTATTGGAGAGCGGAGCTTTGGTTCTTAGTGATCGTGGGATATGC 3359
Db       1750 GTCAGCAGGGAGTGGACCTTGGAGGCTGGGGCCCTGGTTCTGGCTGACCGAGGAGTGTGT 1809

Qy       3360 TGTATCGATGAGTTCGACAAAATGTCTGATAATGCCCGAAGCATGCTTCATGAGGTAATG 3419
Db       1810 CTCATTGATGAATTTGACAAGATGAATGACCAGGACAGAACCAGCATCCATGAGGCCATG 1869

Qy       3420 GAGCAACAAACGGTATCTGTAGCCAAAGGGGGTATCATTGCCTCGCTGAACGCTCGGACG 3479
Db       1870 GAGCAACAGAGCATCTCCATCTCGAAGGCTGGCATCGTCACCTCCCTGCAGGCTCGCTGC 1929

Qy       3480 TCTGTCCTTGCATGTGCAAATCCTAGTGGGTCCCGATACAATGCGCGCCTTTCTGTGATT 3539
Db       1930 ACGGTCATTGCTGCCGCCAACCCCATAGGAGGGCGCTACGACCCCTCGCTGACTTTCTCT 1989

Qy       3540 GATAACATCCAGCTTCCTCCAACTCTACTTTCTAGATTTGATTTAATTTACTTAATGCTC 3599
Db       1990 GAGAACGTGGACCTCACAGAGCCCATCATCTCACGCTTTGACATCCTGTGTGTGGTGAGG 2049

Qy       3600 GACAAACCAGACGAGCAAAACGATCGTCGTCTCGCCAGGCATCTCGTGGCTTTACACTAT 3659
Db       2050 GACACCGTGGACCCAGTCCAGGACGAGATGCTGGCCCGCTTCGTGGTGGGCAGCCACGTC 2109

Qy       3660 GAAAACTATGAAGTTTCAAAGCAGGACG 3687
Db       2110 AGACACCACCCCAGCAACAAGGAGGAGG 2137
```

RESULT 6
US-10-507-275-2
; Sequence 2, Application US/10507275
; Publication No. US20050250166A1
; GENERAL INFORMATION:
; APPLICANT: Masai, Hisao
; APPLICANT: Tamai, Katsuyuki
; APPLICANT: Medical and Biological Laboratories Co., Ltd.
; APPLICANT: Japan Science and Technology Agency
; APPLICANT: Ginkgo Biomedical Research Institute Co., Ltd.
; TITLE OF INVENTION: Cdc7-ASK Kinase Complex, Substrates of the Kinase Complex,
; TITLE OF INVENTION: Specific Antibodies to the Substrates, and Screening Methods
; TITLE OF INVENTION: Using the Same to Screen for Compounds Comprising Cdc7-ASK
; TITLE OF INVENTION: Kinase Inhititory Ability
; FILE REFERENCE: 082368-001100US
; CURRENT APPLICATION NUMBER: US/10/507,275
; CURRENT FILING DATE: 2004-09-09
; PRIOR APPLICATION NUMBER: JP 2002-067702
; PRIOR FILING DATE: 2002-03-12
; PRIOR APPLICATION NUMBER: WO PCT/JP03/02918
; PRIOR FILING DATE: 2003-03-12
; NUMBER OF SEQ ID NOS: 21
; SOFTWARE: PatentIn Ver. 2.1
; SEQ ID NO 2
;   LENGTH: 3379
;   TYPE: DNA
;   ORGANISM: Homo sapiens
;   FEATURE:
;   NAME/KEY: CDS
;   LOCATION: (31)..(2709)
US-10-507-275-2

```
  Query Match             3.6%;  Score 157.6;  DB 6;  Length 3379;
  Best Local Similarity   50.7%;  Pred. No. 8.7e-41;
  Matches  379;  Conservative    0;  Mismatches  369;  Indels    0;  Gaps    0;

Qy       2940 GCTGGGTATCATGAAAGTGATACTTCAGAAGCTGCTAATGAAGCAAAGATTCAAAAACTT 2999
Db       1330 GCCAAGAAGGACAACAAGGTTGCTGTAGGGGAACTGACCGATGAAGATGTGAAGATGATC 1389
```

```
Qy       3000 AAAGAGCTGTCCAAGCTCCCGGGCATTTATGATAGACTTTCAAGGTCGCTGGCTCCAAGC 3059
Db       1390 ACTAGCCTCTCCAAGGATCAGCAGATCGGAGAGAAGATCTTTGCCAGCATTGCTCCTTCC 1449

Qy       3060 ATTTGGGAGCTTGAAGATATTAAAAAGGGTCTTCTTTGCCAGCTCTTTGGTGGGAAGGCT 3119
Db       1450 ATCTATGGTCATGAAGACATCAAGAGAGGCCCTGCTCTGGCCCTGTTCGGAGGGGAGCCC 1509

Qy       3120 AAGAAAATTCCATCTGGAGCATCTTTCCGAGGTGACATCAATGTTTTACTTGTTGGGGAC 3179
Db       1510 AAAAACCCAGGTGGCAAGCACAAGGTACGTGGTGATATCAACGTGCTCTTGTGCGGAGAC 1569

Qy       3180 CCTGGTACCAGTAAATCTCAGCTGCTTCAGTATGTGCACAAGATAGCTCCTCGTGGAATC 3239
Db       1570 CCTGGCACAGCGAAGTCGCAGTTTCTCAAGTATATTGAGAAAGTGTCCAGCCGAGCCATC 1629

Qy       3240 TACACTAGTGGGCGAGGAAGTTCGGCGGTTGGGCTGACAGCGTATGTAACGAAGGATCCA 3299
Db       1630 TTCACCACTGGCCAGGGGGCGTCGGCTGTGGCCGTCACGGCGTATGTCCAGCGGCACCCT 1689

Qy       3300 GAAACTCGAGAGACGGTATTGGAGAGCGGAGCTTTGGTTCTTAGTGATCGTGGGATATGC 3359
Db       1690 GTCAGCAGGGAGTGGACCTTGGAGGCTGGGGCCCTGGTTCTGGCTGACCGAGGAGTGTGT 1749

Qy       3360 TGTATCGATGAGTTCGACAAAATGTCTGATAATGCCCGAAGCATGCTTCATGAGGTAATG 3419
Db       1750 CTCATTGATGAATTTGACAAGATGAATGACCAGGACAGAACCAGCATCCATGAGGCCATG 1809

Qy       3420 GAGCAACAAACGGTATCTGTAGCCAAAGGGGGTATCATTGCCTCGCTGAACGCTCGGACG 3479
Db       1810 GAGCAACAGAGCATCTCCATCTCGAAGGCTGGCATCGTCACCTCCCTGCAGGCTCGCTGC 1869

Qy       3480 TCTGTCCTTGCATGTGCAAATCCTAGTGGGTCCCGATACAATGCGCGCCTTTCTGTGATT 3539
Db       1870 ACGGTCATTGCTGCCGCCAACCCCATAGGAGGGCGCTACGACCCCTCGCTGACTTTCTCT 1929

Qy       3540 GATAACATCCAGCTTCCTCCAACTCTACTTTCTAGATTTGATTTAATTTACTTAATGCTC 3599
Db       1930 GAGAACGTGGACCTCACAGAGCCCATCATCTCACGCTTTGACATCCTGTGTGTGGTGAGG 1989

Qy       3600 GACAAACCAGACGAGCAAAACGATCGTCGTCTCGCCAGGCATCTCGTGGCTTTACACTAT 3659
Db       1990 GACACCGTGGACCCAGTCCAGGACGAGATGCTGGCCCGCTTCGTGGTGGGCAGCCACGTC 2049

Qy       3660 GAAAACTATGAAGTTTCAAAGCAGGACG 3687
Db       2050 AGACACCACCCCAGCAACAAGGAGGAGG 2077
```

RESULT 7
US-11-108-172-158
; Sequence 158, Application US/11108172
; Publication No. US20050260177A1
; GENERAL INFORMATION:
; APPLICANT: Xu, Jiangchun
; APPLICANT: Lodes, Michael J.
; APPLICANT: Secrist, Heather
; APPLICANT: Benson, Darin R.
; APPLICANT: Meagher, Madeleine Joy
; APPLICANT: Stolk, John A.
; APPLICANT: Wang, Tongtong
; APPLICANT: Jiang, Yuqiu
; APPLICANT: Smith, Carole L.
; APPLICANT: King, Gordon E.
; APPLICANT: Wang, Aijun
; APPLICANT: Clapper, Jonathan D.
; APPLICANT: Skeiky, Yasir A. W.
; APPLICANT: Fanger, Gary R.
; APPLICANT: Vedvick Thomas S.
; APPLICANT: Carter, Darrick
; TITLE OF INVENTION: COMPOUNDS FOR IMMUNOTHERAPY AND DIAGNOSIS
; TITLE OF INVENTION: OF COLON CANCER AND METHODS FOR THEIR USE
; FILE REFERENCE: 210121.471C15
; CURRENT APPLICATION NUMBER: US/11/108,172
; CURRENT FILING DATE: 2005-04-15
; PRIOR APPLICATION NUMBER: US 10/025,380
; PRIOR FILING DATE: 2001-12-19
; PRIOR APPLICATION NUMBER: US 09/922,217
; PRIOR FILING DATE: 2001-08-03
; PRIOR APPLICATION NUMBER: US 09/833,263
; PRIOR FILING DATE: 2001-04-10
; PRIOR APPLICATION NUMBER: US 09/649,811
; PRIOR FILING DATE: 2000-08-28
; PRIOR APPLICATION NUMBER: US 09/609,448
; PRIOR FILING DATE: 2000-06-29
; PRIOR APPLICATION NUMBER: US 09/575,251
; PRIOR FILING DATE: 2000-05-19
; PRIOR APPLICATION NUMBER: US 09/519,444
; PRIOR FILING DATE: 2000-03-06
; PRIOR APPLICATION NUMBER: US 09/504,629
; PRIOR FILING DATE: 2000-02-15
; PRIOR APPLICATION NUMBER: US 09/480,321
; PRIOR FILING DATE: 2000-01-10
; PRIOR APPLICATION NUMBER: US 09/476,296
; PRIOR FILING DATE: 1999-12-30
; Remaining Prior Application data removed - See File Wrapper or PALM.
; NUMBER OF SEQ ID NOS: 1130
; SOFTWARE: FastSEQ for Windows Version 4.0
; SEQ ID NO 158
; LENGTH: 507
; TYPE: DNA
; ORGANISM: Homo sapiens
; FEATURE:
; NAME/KEY: misc_feature
; LOCATION: 27, 428, 448, 449, 456, 462, 490, 492, 497, 498, 502, 503
; OTHER INFORMATION: n = A,T,C or G
US-11-108-172-158

Query Match 2.8%; Score 123.2; DB 7; Length 507;
Best Local Similarity 61.8%; Pred. No. 4.4e-30;
Matches 207; Conservative 0; Mismatches 122; Indels 6; Gaps 1;

```
Qy       2992 AAAAACTTAAAGAGCTGTCCAAGCTCCCGGGCATTTATGATAGACTTTCAAGGTCGCTGG 3051
Db        167 AATTGCTTAAGGAACTTTCCAGGAAACCAGACATTTATGAGAGGCTTGCTTCAGCCTTGG 226

Qy       3052 CTCCAAGCATTTGGGAGCTTGAAGATATTAAAAAGGGTCTTCTTTGCCAGCTCTTTGGTG 3111
Db        227 CTCCAAGCATTTATGAACATGAAGATATAAAGAAGGGAATTTTGCTTCAGCTCTTTGGCG 286

Qy       3112 GGAAGGCTAAGAAAATT------CCATCTGGAGCATCTTTCCGAGGTGACATCAATGTTT 3165
Db        287 GGACAAGGAAGGATTTTAGTCACACTGGAAGGGGCAAATTTCGGGCTGAGATCAACATCT 346

Qy       3166 TACTTGTTGGGGACCCTGGTACCAGTAAATCTCAGCTGCTTCAGTATGTGCACAAGATAG 3225
Db        347 TGCTGTGTGGCGACCCTGGTACCAGCAAGTCCCAGCTGCTGCAGTACGTGTACAACCTCG 406

Qy       3226 CTCCTCGTGGAATCTACACTAGTGGGCGAGGAAGTTCGGCGGTTGGGCTGACAGCGTATG 3285
Db        407 TCCCCAGGGGCCAGTACACGTNTGGGAAGGGCTCCAGTGCANNTGGCCTNACTGCNTACG 466

Qy       3286 TAACGAAGGATCCAGAAACTCGAGAGACGGTATTG 3320
Db        467 TAATGAAAGACCCTGAGACAAGGNANCTGGNNCTG 501
```

RESULT 8
US-10-750-185-45476
; Sequence 45476, Application US/10750185
; Publication No. US20050260603A1
; GENERAL INFORMATION:
; APPLICANT: MMI GENOMICS, INC.
; APPLICANT: DeNISE, Sue K.
; APPLICANT: KERR, Richard

```
;  APPLICANT:  ROSENFELD, David
;  APPLICANT:  HOLM, Tom
;  APPLICANT:  BATES, Stephen
;  APPLICANT:  FANTIN, Dennis
;  TITLE OF INVENTION: COMPOSITIONS FOR INFERRING BOVINE TRAITS
;  FILE REFERENCE: MMI1100-2
;  CURRENT APPLICATION NUMBER: US/10/750,185
;  CURRENT FILING DATE:  2003-12-31
;  PRIOR APPLICATION NUMBER: US 60/437,482
;  PRIOR FILING DATE: 2002-12-31
;  NUMBER OF SEQ ID NOS: 64922
;  SOFTWARE: PatentIN version 3.1
; SEQ ID NO 45476
;   LENGTH: 1851
;   TYPE: DNA
;   ORGANISM: Bovine  19866880565574
US-10-750-185-45476

  Query Match             1.3%;  Score 58.2;  DB 6;  Length 1851;
  Best Local Similarity   59.3%;  Pred. No. 2.8e-08;
  Matches   99;  Conservative    0;  Mismatches   68;  Indels    0;  Gaps    0;

Qy       3285 GTAACGAAGGATCCAGAAACTCGAGAGACGGTATTGGAGAGCGGAGCTTTGGTTCTTAGT 3344
              |||||      |       |||  |  ||||    |    ||||||    ||||| || || ||
Db       1381 GTAACTGCTGTAAAAGACTCAGGAGAATGGAATTTGGAGGCGGGAGCATTAGTCCTCGCA 1440

Qy       3345 GATCGTGGGATATGCTGTATCGATGAGTTCGACAAAATGTCTGATAATGCCCGAAGCATG 3404
              |||  |||   | ||||||||||||| ||   | |    |      ||  |||  | |||   |
Db       1441 GATGCTGGCCTCTGCTGTATCGATGAATTTAATAGCCTCAAAGAGCATGACAGAACTAGT 1500

Qy       3405 CTTCATGAGGTAATGGAGCAACAAACGGTATCTGTAGCCAAAGGGGG 3451
               | || || |  |||||||| |||||   |      ||| || || |  ||
Db       1501 ATCCACGAAGCAATGGAGCAGCAAACCATCAGTGTTGCTAAGGCTGG 1547

RESULT 9
US-10-750-185-55669
; Sequence 55669, Application US/10750185
; Publication No. US20050260603A1
; GENERAL INFORMATION:
;  APPLICANT: MMI GENOMICS, INC.
;  APPLICANT:  DeNISE, Sue K.
;  APPLICANT:  KERR, Richard
;  APPLICANT:  ROSENFELD, David
;  APPLICANT:  HOLM, Tom
;  APPLICANT:  BATES, Stephen
;  APPLICANT:  FANTIN, Dennis
;  TITLE OF INVENTION: COMPOSITIONS FOR INFERRING BOVINE TRAITS
;  FILE REFERENCE: MMI1100-2
;  CURRENT APPLICATION NUMBER: US/10/750,185
;  CURRENT FILING DATE:  2003-12-31
;  PRIOR APPLICATION NUMBER: US 60/437,482
;  PRIOR FILING DATE: 2002-12-31
;  NUMBER OF SEQ ID NOS: 64922
;  SOFTWARE: PatentIN version 3.1
; SEQ ID NO 55669
;   LENGTH: 1312
;   TYPE: DNA
;   ORGANISM: Bovine  19866880618533
US-10-750-185-55669

  Query Match             1.3%;  Score 54.8;  DB 6;  Length 1312;
  Best Local Similarity   61.0%;  Pred. No. 2.8e-07;
  Matches   89;  Conservative    0;  Mismatches   57;  Indels    0;  Gaps    0;

Qy       3238 TCTACACTAGTGGGCGAGGAAGTTCGGCGGTTGGGCTGACAGCGTATGTAACGAAGGATC 3297
              | |||||    ||||   ||| ||        || |  |||||||||||   || |  || ||| |
Db        841 TGTACACGTCTGGGAAAGGCAGCAGCGCCGCCGGGCTGACAGCCTCGGTGATGAGGGACC 900

Qy       3298 CAGAAACTCGAGAGACGGTATTGGAGAGCGGAGCTTTGGTTCTTAGTGATCGTGGGATAT 3357
              |      | ||  |          |  ||||| |||| ||   |||| ||      ||  | ||   |
Db        901 CCTCGTCCCGGAACTTCATCATGGAGGGCGGGGCCATGGTCCTGGCCGACGGAGGTGTCG 960

Qy       3358 GCTGTATCGATGAGTTCGACAAAATG 3383
                |||||| |||||||| |||||   ||
Db        961 TCTGTATTGATGAGTTTGACAAGGTG 986

RESULT 10
US-10-750-185-27656
; Sequence 27656, Application US/10750185
; Publication No. US20050260603A1
; GENERAL INFORMATION:
;  APPLICANT: MMI GENOMICS, INC.
;  APPLICANT:  DeNISE, Sue K.
;  APPLICANT:  KERR, Richard
;  APPLICANT:  ROSENFELD, David
;  APPLICANT:  HOLM, Tom
;  APPLICANT:  BATES, Stephen
;  APPLICANT:  FANTIN, Dennis
;  TITLE OF INVENTION: COMPOSITIONS FOR INFERRING BOVINE TRAITS
;  FILE REFERENCE: MMI1100-2
;  CURRENT APPLICATION NUMBER: US/10/750,185
;  CURRENT FILING DATE:  2003-12-31
;  PRIOR APPLICATION NUMBER: US 60/437,482
;  PRIOR FILING DATE: 2002-12-31
;  NUMBER OF SEQ ID NOS: 64922
;  SOFTWARE: PatentIN version 3.1
; SEQ ID NO 27656
;   LENGTH: 876
;   TYPE: DNA
;   ORGANISM: Bovine  19866881490050
US-10-750-185-27656

  Query Match             0.9%;  Score 39.6;  DB 6;  Length 876;
  Best Local Similarity   58.5%;  Pred. No. 0.021;
  Matches   69;  Conservative    0;  Mismatches   49;  Indels    0;  Gaps    0;

Qy       1260 CTTCAAGAAACTCCAGATGCCATTCCTGAAGGAGAGACTCCACACACAGTCAGCATGTGT 1319
              || || ||  |   | || || ||| || |   ||   |||| || ||||| |||   | | | |
Db        508 CTCCAGGAGTCCCCTGAAGACATGCCCGCGGGGCAGACGCCGCACACGGTCGTCCTCTTT 567

Qy       1320 TTATACAACACTATGGTTGATGCTGTGAAGCCTGGAGATCGTATTGAGGTAACAGGAG 1377
                  ||||     | || ||    ||  ||||||| ||    |    |  || ||||| |
Db        568 GCTCACAATGACCTCGTGGACAAGGTCCAGCCTGGGGACAGAGTGCACGTCACAGGTG 625

RESULT 11
US-10-821-234-393
; Sequence 393, Application US/10821234
; Publication No. US20050255114A1
; GENERAL INFORMATION:
;  APPLICANT: Labat, Ivan
;  APPLICANT:  Stache-Crain, Birgit
;  APPLICANT:  Andarmani, Susan
;  APPLICANT:  Tang, Y. Tom
;  TITLE OF INVENTION: Methods for Diagnosis and Treatment of Preeclampsia
;  FILE REFERENCE: 821A
;  CURRENT APPLICATION NUMBER: US/10/821,234
;  CURRENT FILING DATE:  2004-04-07
;  PRIOR APPLICATION NUMBER: US 60/462,047
;  PRIOR FILING DATE: 2003-04-07
;  NUMBER OF SEQ ID NOS: 1704
;  SOFTWARE: pt_SEQ_genes Version 1.0
; SEQ ID NO 393
;   LENGTH: 449
;   TYPE: DNA
;   ORGANISM: Homo sapiens
;   FEATURE:
;   NAME/KEY: misc_feature
;   LOCATION: (1)...(449)
;   OTHER INFORMATION: n = a,t,c or g
US-10-821-234-393

  Query Match             0.9%;  Score 37.6;  DB 6;  Length 449;
  Best Local Similarity   54.3%;  Pred. No. 0.058;
```

```
  Matches   76;  Conservative    0;  Mismatches   64;  Indels    0;  Gaps    0;

Qy       4115 TGGCTTGAAGACCAGTCAGCTTCTTGAGGATATCCGGAGCCAAAGCAGTGTGGACGTTAG 4174

Db         65 TGTCTCCAGAGCTGTCCAGGAATTTGGTCTTGCCCGGTTCAAAAGCAACGTGACCAAGAC 124

Qy       4175 TTTGCAGGATATTAAAAATGCTCTGGGTAGCCTCCAAGGAGAAGGCTTTCTTACTGTCCA 4234

Db        125 TATGAAGGGTTTTGAATATATCTTGGCTAAGCTGCAAGGCGAGGCCCCTTCCAAAACACT 184

Qy       4235 TGGTGACATAGTCAAGAGAG 4254

Db        185 TGTTGAGACAGCCAAGGGAG 204


RESULT 12
US-10-821-234-409/c
; Sequence 409, Application US/10821234
; Publication No. US20050255114A1
; GENERAL INFORMATION:
;  APPLICANT: Labat, Ivan
;  APPLICANT:  Stache-Crain, Birgit
;  APPLICANT:  Andarmani, Susan
;  APPLICANT:  Tang, Y. Tom
;  TITLE OF INVENTION: Methods for Diagnosis and Treatment of Preeclampsia
;  FILE REFERENCE: 821A
;  CURRENT APPLICATION NUMBER: US/10/821,234
;  CURRENT FILING DATE:  2004-04-07
;  PRIOR APPLICATION NUMBER: US 60/462,047
;  PRIOR FILING DATE: 2003-04-07
;  NUMBER OF SEQ ID NOS: 1704
;  SOFTWARE: pt_SEQ_genes Version 1.0
; SEQ ID NO 409
;   LENGTH: 978
;   TYPE: DNA
;   ORGANISM: Homo sapiens
US-10-821-234-409

  Query Match             0.8%;  Score 35.6;  DB 6;  Length 978;
  Best Local Similarity   55.7%;  Pred. No. 0.47;
  Matches   68;  Conservative    0;  Mismatches   54;  Indels    0;  Gaps    0;

Qy       2994 AAACTTAAAGAGCTGTCCAAGCTCCCGGGCATTTATGATAGACTTTCAAGGTCGCTGGCT 3053

Db        519 ATACTGGTTGAGCTCTCCACTCTCATGGCCAGCGATGATGCACTCCCCCAGGGGTCCCCA 460

Qy       3054 CCAAGCATTTGGGAGCTTGAAGATATTAAAAAGGGTCTTCTTTGCCAGCTCTTTGGTGGG 3113

Db        459 AACAGCACTGGTGATTTTAGAGTCATTGCAAGGGATCTTCATGTAGGGCTCATTGTTGTC 400

Qy       3114 AA 3115

Db        399 AA 398


RESULT 13
US-10-793-626-4034/c
; Sequence 4034, Application US/10793626
; Publication No. US20050255478A1
; GENERAL INFORMATION:
;  APPLICANT: KIMMERLY, WILLIAM JOHN
;  TITLE OF INVENTION: STAPHYLOCOCCUS EPIDERMIDIS NUCLEIC ACIDS AND PROTEINS
;  FILE REFERENCE: PU3480US
;  CURRENT APPLICATION NUMBER: US/10/793,626
;  CURRENT FILING DATE:  2004-03-04
;  PRIOR APPLICATION NUMBER: 60/164,258
;  PRIOR FILING DATE: 1999-11-09
;  NUMBER OF SEQ ID NOS: 4472
;  SOFTWARE: PatentIn Ver. 2.1
; SEQ ID NO 4034
;   LENGTH: 3150
;   TYPE: DNA
;   ORGANISM: Artificial Sequence
;   FEATURE:
;   OTHER INFORMATION: Description of Artificial Sequence: synthetic
;   OTHER INFORMATION: nucleic acid sequence
US-10-793-626-4034

  Query Match             0.8%;  Score 35.4;  DB 6;  Length 3150;
  Best Local Similarity   51.6%;  Pred. No. 1.3;
  Matches   81;  Conservative    0;  Mismatches   76;  Indels    0;  Gaps    0;

Qy       2886 GGTCGACTGCAAACTGAAGATCCTATGGAGATGGATAAGGAGAATGATATGTATGCTGGG 2945

Db       2236 GGTAGAAATTATGAAGATCATGCTAGTGAATTAAATCATGAAGTTGATAGTTTATACGTA 2177

Qy       2946 TATCATGAAAGTGATACTTCAGAAGCTGCTAATGAAGCAAAGATTCAAAAACTTAAAGAG 3005

Db       2176 TTTACTAAAGCGGCTTCATCTTTAACTGGTGATGAAGCAACAATTCCTAATCATAAAGAT 2117

Qy       3006 CTGTCCAAGCTCCCGGGCATTTATGATAGACTTTCAA 3042

Db       2116 ATTACAGAACAATTGGACTATGAAGGCGAACTTGGAA 2080


RESULT 14
US-10-793-626-4115
; Sequence 4115, Application US/10793626
; Publication No. US20050255478A1
; GENERAL INFORMATION:
;  APPLICANT: KIMMERLY, WILLIAM JOHN
;  TITLE OF INVENTION: STAPHYLOCOCCUS EPIDERMIDIS NUCLEIC ACIDS AND PROTEINS
;  FILE REFERENCE: PU3480US
;  CURRENT APPLICATION NUMBER: US/10/793,626
;  CURRENT FILING DATE:  2004-03-04
;  PRIOR APPLICATION NUMBER: 60/164,258
;  PRIOR FILING DATE: 1999-11-09
;  NUMBER OF SEQ ID NOS: 4472
;  SOFTWARE: PatentIn Ver. 2.1
; SEQ ID NO 4115
;   LENGTH: 4231
;   TYPE: DNA
;   ORGANISM: Artificial Sequence
;   FEATURE:
;   OTHER INFORMATION: Description of Artificial Sequence: synthetic
;   OTHER INFORMATION: nucleic acid sequence
US-10-793-626-4115

  Query Match             0.8%;  Score 35.4;  DB 6;  Length 4231;
  Best Local Similarity   51.6%;  Pred. No. 1.6;
  Matches   81;  Conservative    0;  Mismatches   76;  Indels    0;  Gaps    0;

Qy       2886 GGTCGACTGCAAACTGAAGATCCTATGGAGATGGATAAGGAGAATGATATGTATGCTGGG 2945

Db       1768 GGTAGAAATTATGAAGATCATGCTAGTGAATTAAATCATGAAGTTGATAGTTTATACGTA 1827

Qy       2946 TATCATGAAAGTGATACTTCAGAAGCTGCTAATGAAGCAAAGATTCAAAAACTTAAAGAG 3005

Db       1828 TTTACTAAAGCGGCTTCATCTTTAACTGGTGATGAAGCAACAATTCCTAATCATAAAGAT 1887

Qy       3006 CTGTCCAAGCTCCCGGGCATTTATGATAGACTTTCAA 3042

Db       1888 ATTACAGAACAATTGGACTATGAAGGCGAACTTGGAA 1924


RESULT 15
US-10-793-626-3317
; Sequence 3317, Application US/10793626
; Publication No. US20050255478A1
; GENERAL INFORMATION:
;  APPLICANT: KIMMERLY, WILLIAM JOHN
;  TITLE OF INVENTION: STAPHYLOCOCCUS EPIDERMIDIS NUCLEIC ACIDS AND PROTEINS
;  FILE REFERENCE: PU3480US
;  CURRENT APPLICATION NUMBER: US/10/793,626
;  CURRENT FILING DATE:  2004-03-04
;  PRIOR APPLICATION NUMBER: 60/164,258
```

```
;  PRIOR FILING DATE: 1999-11-09
;  NUMBER OF SEQ ID NOS: 4472
;  SOFTWARE: PatentIn Ver. 2.1
; SEQ ID NO 3317
;   LENGTH: 1263
;   TYPE: DNA
;   ORGANISM: Artificial Sequence
;   FEATURE:
;   OTHER INFORMATION: Description of Artificial Sequence: synthetic
;   OTHER INFORMATION: nucleic acid sequence
US-10-793-626-3317

  Query Match             0.8%;  Score 34.2;  DB 6;  Length 1263;
  Best Local Similarity   55.5%;  Pred. No. 1.6;
  Matches   66;  Conservative    0;  Mismatches   53;  Indels    0;  Gaps    0;

Qy       2925 GAGAATGATATGTATGCTGGGTATCATGAAAGTGATACTTCAGAAGCTGCTAATGAAGCA 2984
              |||||| |||    || |    |    ||||   || | ||||||| |  |||  ||||
Db         28 GAGAATTATAATATTGTTACAGAAATAGAAAAATATGCCTCAGAAGATCATAAAAAAGCC 87

Qy       2985 AAGATTCAAAAACTTAAAGAGCTGTCCAAGCTCCCGGGCATTTATGATAGACTTTCAAG 3043
              |  ||| | ||   ||| ||||     ||  |    |  | |||| |    |||   ||
Db         88 ATTATTTACAAGGATAACGAGCATGAAAATATTTCTGTAAGTTATAAAGAACTTATCAG 146


Search completed: December  6, 2005, 09:59:11
Job time : 398 secs
```

```
                    GenCore version 5.1.6
          Copyright (c) 1993 - 2005  Compugen Ltd.

OM nucleic - nucleic search, using sw model

Run on:         December  6, 2005, 01:29:25 ; Search time 10575 Seconds
                                           (without alignments)
                                           19236.902 Million cell updates/sec

Title:          US-10-768-511-5
Perfect score:  4348
Sequence:       1 atggcgcgccgcactcacgt..........gcgcaattcccagagctcgc 4348

Scoring table:  IDENTITY_NUC
                Gapop 10.0 , Gapext 1.0

Searched:       41078325 seqs, 23393541228 residues

Total number of hits satisfying chosen parameters:      82156650

Minimum DB seq length: 0
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%
                 Maximum Match 100%
                 Listing first 45 summaries

Database :       EST:*
                1:  gb_est1:*
                2:  gb_est2:*
                3:  gb_est3:*
                4:  gb_htc:*
                5:  gb_est4:*
                6:  gb_est5:*
                7:  gb_est6:*
                8:  gb_est7:*
                9:  gb_gss1:*
                10:  gb_gss2:*
                11:  gb_gss3:*

      Pred. No. is the number of results predicted by chance to have a
      score greater than or equal to the total score of the result being printed,
      and is derived by analysis of the total score distribution.

                              SUMMARIES
                          %
Result                  Query
   No.    Score     Match Length DB ID                        Description
---------------------------------------------------------------------------
     1    871.2      20.0   2699  4  CNS0A83A                 BX820670 Arabidops
     2      777      17.9   3168 10  CL959068                 CL959068 OsIFCC002
c    3      478      11.0    879  8  DR495517                 DR495517 WS0298.BR
     4    469.6      10.8    881  7  CO123277                 CO123277 GR__Eb05F
     5    456.4      10.5   1156  7  CK210237                 CK210237 FGAS02203
     6    439.4      10.1    907  8  DR503163                 DR503163 WS0298.B2
     7    434.2      10.0    770  2  BF051102                 BF051102 EST436277
     8      433      10.0    788  3  BM412106                 BM412106 EST586433
     9    429.2       9.9   2522 10  AY408033                 AY408033 Homo sapi
    10      428       9.8    775  3  BM410852                 BM410852 EST585179
    11    408.4       9.4   2357 10  AY408034                 AY408034 Pan trogl
    12    404.2       9.3    798  6  CF449139                 CF449139 EST685484
    13    394.4       9.1    709  8  DR917805                 DR917805 EST110934
    14    392.8       9.0   3285  4  AK089999                 AK089999 Mus muscu
    15    392.8       9.0   3301  4  AK088796                 AK088796 Mus muscu
    16    392.8       9.0   3545  4  AK036653                 AK036653 Mus muscu
    17    392.6       9.0   2519 10  AY408035                 AY408035 Mus muscu
    18    391.2       9.0   3316  4  AK011743                 AK011743 Mus muscu
    19    387.4       8.9    750  6  CA258367                 CA258367 SCCCRT300
    20    379.6       8.7    706  3  BJ464173                 BJ464173 BJ464173
    21    378.6       8.7    765  6  CF442356                 CF442356 EST678701
c   22      377       8.7   1082  5  BX836107                 BX836107 BX836107
    23    375.2       8.6    727  7  CK072837                 CK072837 68036rsic
    24    374.8       8.6    755  8  DN232047                 DN232047 MEST1067_
    25    364.2       8.4    615  5  BQ465527                 BQ465527 HU03012r
    26    363.2       8.4    698  7  CV461490                 CV461490 aof02-11m
    27    358.4       8.2    717  6  CA501849                 CA501849 WHE4039_E
    28    357.8       8.2   1607  4  CR598188                 CR598188 full-leng
    29    357.8       8.2   1833  4  CR606514                 CR606514 full-leng
    30      356       8.2    910  8  CX099871                 CX099871 RECm0634
    31    352.4       8.1    619  3  BQ048103                 BQ048103 1091002G0
    32    349.2       8.0    638  5  BU989891                 BU989891 HF22P19r
    33    348.8       8.0    668  6  CA258647                 CA258647 SCCCRT301
    34      342       7.9    644  7  CV008987                 CV008987 atr01-32m
    35    339.8       7.8    584  3  BQ048191                 BQ048191 1091004C0
    36    339.8       7.8    673  7  CV055224                 CV055224 BNEL118b9
    37    336.6       7.7    642  6  CD475195                 CD475195 nad03-20m
    38    330.8       7.6    560  6  CD480657                 CD480657 eca01-6ms
    39    329.8       7.6    823  8  CV881487                 CV881487 Mdfrt3117
    40      329       7.6    593  5  BQ459696                 BQ459696 HA08F14r
    41    328.6       7.6    725  6  CA258377                 CA258377 SCCCRT300
    42    327.8       7.5    632  8  CV794540                 CV794540 Mdfrt3114
    43      326       7.5    676  3  BP955291                 BP955291 BP955291
    44    322.6       7.4    571  3  BQ048021                 BQ048021 1091001E1
    45    319.2       7.3    613  6  CA827375                 CA827375 1114013G0

                              ALIGNMENTS

RESULT 1
CNS0A83A
LOCUS       CNS0A83A               2699 bp    mRNA    linear   HTC 06-FEB-2004
DEFINITION  Arabidopsis thaliana Full-length cDNA Complete sequence from clone
            GSLTPGH58ZF05 of Hormone Treated Callus of strain col-0 of
            Arabidopsis thaliana (thale cress).
ACCESSION   BX820670
VERSION     BX820670.1  GI:42467030
KEYWORDS    HTC; GSLT_cDNA.
SOURCE      Arabidopsis thaliana (thale cress)
  ORGANISM  Arabidopsis thaliana
            Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
            Spermatophyta; Magnoliophyta; eudicotyledons; core eudicotyledons;
            rosids; eurosids II; Brassicales; Brassicaceae; Arabidopsis.
REFERENCE   1  (bases 1 to 2699)
  AUTHORS   Castelli,V., Aury,J.M., Jaillon,O., Wincker,P., Clepet,C.,
            Menard,M., Cruaud,C., Quetier,F., Scarpelli,C., Schachter,V.,
            Temple,G., Caboche,M., Weissenbach,J. and Salanoubat,M.
  TITLE     Whole Genome Sequence Comparisons and 'Full-Length' cDNA Sequences:
            A Combined Approach to Evaluate and Improve Arabidopsis Genome
            Annotation
  JOURNAL   Unpublished
REFERENCE   2  (bases 1 to 2699)
  AUTHORS   Genoscope.
  TITLE     Direct Submission
  JOURNAL   Submitted (18-NOV-2003) Genoscope - Centre National de Sequencage :
            BP 191 91006 EVRY cedex - FRANCE (E-mail : seqref@genoscope.cns.fr
            - Web : www.genoscope.cns.fr)
COMMENT     The sequences are based on single pass reads.
            Life Technologies (a division of Invitrogen) members carried out
            full-length librairies construction : Temple G.
            Genoscope members carried out sequencing and annotation : Castelli
            V., Aury J.M., Jaillon O., Wincker P., Menard M., Cruaud C.,
            Schachter V., Weissenbach J., Salanoubat M.
            URGV INRA : Clepet C., Caboche M.
            Annotation is based on the June 2003 version of the Arabidopsis
            genome released by MIPS (Munich Information center for Protein
            Sequences). 5 prime and 3 prime are assembled with Phrap.
            http://www.genoscope.cns.fr/externe/sequences/Banque_Projet_EF/Full
            _length
            http://www.genoscope.cns.fr/cgi-bin/ggb/ggb?source=Arabidopsis.
FEATURES             Location/Qualifiers
     source          1..2699
                     /organism="Arabidopsis thaliana"
                     /mol_type="mRNA"
```

```
                    /db_xref="taxon:3702"
                    /clone="GSLTPGH58ZF05"
                    /tissue_type="Hormone Treated Callus"
                    /ecotype="Col-0"
                    /plasmid="pCMVSPORT_6"
     gene           complement(1..2699)
                    /gene="At2g16440"
ORIGIN

  Query Match             20.0%;  Score 871.2;  DB 4;  Length 2699;
  Best Local Similarity   64.5%;  Pred. No. 3e-257;
  Matches 1346;  Conservative    0;  Mismatches  718;  Indels   24;  Gaps    2;

Qy       2178 AAGTACATCCAGATCATAGAGGAGACTGTGGAGCGTGAGGAGGATACTCTAAATATCGAC 2237
Db        478 AAGTATATGGTTTCGATAAGGAAGGTGATTGAGATTGAAGGAGAATGGATTGATGTTGAT 537

Qy       2238 ATGTCAGACATTTATGACCATGATCCTGATCTATACGCAAAAATTGTTCGATACCCACTC 2297
Db        538 GCGTTTGATGTGTTTGATTATGATCCTGATTTGTATAACAAGATGGTTCGTTATCCTTTG 597

Qy       2298 GACATCATCCCCCTGTTGGACACTGAGTGTCAGGAAGTTGCTACCTCTTTACTACCAACG 2357
Db        598 GAAGTTCTCGCTATCTTCGATATTGTGCTGATGGATATTGTTTCCACGATTAATCGTTTG 657

Qy       2358 TTTGAGAAGCATATTGAGGCCAGACCTTTCAATCTCAAAGCATCGGTGCACATGCGTGAA 2417
Db        658 TTCGAGAAACATGTTCAAGTTAGGATTTTTAACCTCAGGACATCTACTTCCATGAGAAAT 717

Qy       2418 CTCAACCCTTCAGATATAGACAAATTGGTTTCTGTTAAAGGAATGGTTATCCGGTGCAGT 2477
Db        718 CTCAATCCATCTGATATCGAGAAGATGATCTCTTTGAAAGGTATGATAATTCGGAGTAGC 777

Qy       2478 TCTATCATACCTGAAATTAAGGGGGCCTTCTTCAAATGTTTAGTGTGTGGTCACTCGCCT 2537
Db        778 TCTATCATTCCTGAAATAAGGGAAGCTGTGTTTAGATGCCTTGTTTGTGGTTACTTCTCT 837

Qy       2538 CCGCTAGTTACAGTTGTTAAAGGGCGGGTTGAGGAGCCAACAAGGTGTGAAAAGCCAGAA 2597
Db        838 GACCCCATCATCGTTGATAGAGGGAAAATAAGTGAGCCTCCTACTTGCTTGAAACAAGAG 897

Qy       2598 TGTGCAGCACGGAATGCTATGTCTCTTATTCACAATCGATGCACTTTTGCAAATAAGCAG 2657
Db        898 TGCATGACCAAGAACTCCATGACACTAGTGCACAACCGATGCAGGTTTGCTGATAAGCAG 957

Qy       2658 ATAGTGCGTCTTCAAGAAACTCCAGATGCCATTCCTGAAGGAGAGACTCCACACACAGTC 2717
Db        958 ATTGTGAGGCTTCAGGAAACACCTGATGAGATTCCAGAAGGAGGAACACCTCACACAGTT 1017

Qy       2718 AGCATGTGTTTATACAACACTATGGTTGATGCTGTGAAGCCTGGAGATCGTATTGAGGTA 2777
Db       1018 AGCTTGTTGCTGCATGATAAGCTAGTTGATAATGGAAGCCTGGTGATAGAATTGAGGTC 1077

Qy       2778 ACAGGAGTTTTCAAGGCCATGGCAGTTCGAGTTGGTCCGAATCAACGAACATTACGAGCA 2837
Db       1078 ACCGGAATTTACAGAGCAATGACTGTCCGAGTGGGACCTGCTCACAGGACTGTGAAATCT 1137

Qy       2838 TTGTATAAGACCTACATCGATTGCGTGCACGTCAAGAAGTCTGACAGGGGTCGACTGCAA 2897
Db       1138 GTGTTTAAGACCTATATCGATTGCCTTCACATAAAGAAAGCAAGTAAGTTAAGAATGTCT 1197

Qy       2898 ACTGAAGATCCTATGGAGATGGATAAGGAGAATGATATGTATGCTGGGTATCATGAAAGT 2957
Db       1198 GCTGAGGATCCTATGGATGTTGACAATAGTCTGCGTAGAGTTGATGAAGATGTCGAGTTA 1257

Qy       2958 GATACTTCAGAAGCTGCTAATGAAGCAAAGATTCAAAAACTTAAAGAGCTGTCCAAGCTC 3017
Db       1258 G------------------ATGAGGAGAAGCTGAGGAAGTTTCAGGAACTCTCTAAACAA 1299

Qy       3018 CCGGGCATTTATGATAGACTTTCAAGGTCGCTGGCTCCAAGCATTTGGGAGCTTGAAGAT 3077
Db       1300 CCAGATATATATGAGAGGCTTAGTAGATCACTAGCACCAAACATCTGGGAGTTGGATGAT 1359

Qy       3078 ATTAAAAAGGGTCTTCTTTGCCAGCTCTTTGGTGGGAAGGCTAAGAAAATTCCATCTGGA 3137
Db       1360 GTTAAAAAGGGTCTTCTATGCCAGCTTTTTGGAGGGAATGCTTTGAACTTAGCATCTGGT 1419

Qy       3138 GCATCTTTCCGAGGTGACATCAATGTTTTACTTGTTGGGGACCCTGGTACCAGTAAATCT 3197
Db       1420 GCTAATTTTCGTGGTGACATCAACATCTTACTTGTTGGTGACCCTGGTACGAGCAAGTCC 1479

Qy       3198 CAGCTGCTTCAGTATGTGCACAAGATAGCTCCTCGTGGAATCTACACTAGTGGGCGAGGA 3257
Db       1480 CAGCTGCTCCAGTACATTCATAAGCTTTCACCCCGTGGCATTTACACAAGTGGGCGAGGG 1539

Qy       3258 AGTTCGGCGGTTGGGCTGACAGCGTATGTAACGAAGGATCCAGAAACTCGAGAGACGGTA 3317
Db       1540 AGCTCAGCTGTTGGTTTAACAGCTTATGTAGCTAAAGATCCTGAGACAGGAGAAACTGTT 1599

Qy       3318 TTGGAGAGCGGAGCTTTGGTTCTTAGTGATCGTGGGATATGCTGTATCGATGAGTTCGAC 3377
Db       1600 TTGGAGAGTGGAGCTCTTGTTCTTAGTGACCGAGGTATCTGCTGTATTGATGAATTTGAC 1659

Qy       3378 AAAATGTCTGATAATGCCCGAAGCATGCTTCATGAGGTAATGGAGCAACAAACGGTATCT 3437
Db       1660 AAAATGTCTGACAGTGCGAGGAGCATGCTACACGAGGTGATGGAACAGCAGACTGTTTCA 1719

Qy       3438 GTAGCCAAAGGGGGTATCATTGCCTCGCTGAACGCTCGGACGTCTGTCCTTGCATGTGCA 3497
Db       1720 ATAGCAAAGGCTGGTATTATTGCATCTCTAAATGCCAGAACCTCTGTGTTGGCTTGTGCA 1779

Qy       3498 AATCCTAGTGGGTCCCGATACAATGCGCGCCTTTCTGTGATTGATAACATCCAGCTTCCT 3557
Db       1780 AATCCTAGTGGCTCACGTTATAATCCCCGGCTTTCTGTTATTGAGAATATTCACCTTCCT 1839

Qy       3558 CCAACTCTACTTTCTAGATTTGATTTAATTTACTTAATGCTCGACAAACCAGACGAGCAA 3617
Db       1840 CCAACCTTGCTATCTAGATTCGATTTGATCTACTTGATTCTTGACAAGCCCGATGAGCAG 1899

Qy       3618 AACGATCGTCGTCTCGCCAGGCATCTCGTGGCTTTACACTATGAAAACTATGAAGTTTCA 3677
Db       1900 ACTGACCGAAGGCTTGCAAAGCATATTGTGGCCCTTCACTTTGAGAACGCTGAGAGTGCT 1959

Qy       3678 AAGCAGGACGCCTTAGATCTACAAACACTTACCGCGTATATCACCTATGCTCGTCAGCAT 3737
Db       1960 CAGGAGGAAGCTATAGATATTACTACACTGACAACTTATGTTAGCTATGCCCGCAAGAAC 2019

Qy       3738 GTACATCCTACATTAAGTGATGAAGCTGCTGAAGATTTGATTAATGGCTATGTTGAGATG 3797
Db       2020 ATTCATCCTAAACTGTCAGATGAAGCTGCAGAGGAGTTGACCCGAGGATATGTTGAGCTG 2079

Qy       3798 CGCCAAAAGGGCAACTTTCCTGGAAGCAGTAAAAAGGTGATAACAGCCACACCTCGGCAA 3857
Db       2080 AGAAAAGCAGGGAAGTTTGCTGGCAGTAGCAAAAAGGTCATAACAGCAACTCCTAGGCAG 2139

Qy       3858 CTCGAAAGTATGATTCGTATCAGTGAAGCCCTAGCTCGAATGAGATTTTCTGAAGTGGTA 3917
Db       2140 ATTGAAAGTTTGATCAGACTTAGTGAGGCCTTGGCTCGGATGCGCTTCTCCGAATGGGTT 2199

Qy       3918 GAGAAAGTTGATGCAGCAGAAGCTGTGCGCCTTTTAGACGTCGCTTTGCAGCAATCTGCT 3977
Db       2200 GAAAAGCATGACGTGGACGAGGCATTTCGACTTCTCAGAGTTGCAATGCAGCAATCAGCA 2259

Qy       3978 ACTGATCATGCAACAGGTACGATAGACATGGATCTTATCACGACTGGAGTGTCGGCCAGC 4037
Db       2260 ACGGATCACGCCACAGGGACCATTGACATGGACCTGATAAATACTGGAGTGTCAGCAAGT 2319

Qy       4038 GAGCGTATTCGTCGGGCCAACTTGCTAGCTGCTCTGCGAGAGCTTATAGCAGATAAAATT 4097
Db       2320 GAACGAATGAGGCGGGATACCTTTGCGTCGTCAATTAGAGACATAGCTCTGGAGAAGATG 2379

Qy       4098 TCACCTGGCAGCTCCTC------TGGCTTGAAGACCAGTCAGCTTCTTGAGGATATCCGG 4151
Db       2380 CAAATTGGAGGATCATCAATGCGCTTATCCGAGTTACTTGAAGAACTGAAGAAGCATGGA 2439

Qy       4152 AGCCAAAGCAGTGTGGACGTTAGTTTGCAGGATATTAAAAATGCTCTGGGTAGCCTCCAA 4211
```

```
Db     2440 GGCAACATAAACACCGAAATTCATCTTCATGATGTTAGAAAAGCAGTCGCGACGCTAGCG 2499

Qy     4212 GGAGAAGGCTTTCTTACTGTCCATGGTGACATAGTCAAGAGAGTTTGA 4259

Db     2500 AGTGAAGGATTTCTGGTAGCTGAAGGTGATAGAATCAAGAGAGTATAA 2547


RESULT 2
CL959068
LOCUS       CL959068                3168 bp    DNA     linear   GSS 21-SEP-2004
DEFINITION  OsIFCC002182 Oryza sativa Express Library Oryza sativa (indica
            cultivar-group) genomic, genomic survey sequence.
ACCESSION   CL959068
VERSION     CL959068.1  GI:52372893
KEYWORDS    GSS.
SOURCE      Oryza sativa (indica cultivar-group)
  ORGANISM  Oryza sativa (indica cultivar-group)
            Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
            Spermatophyta; Magnoliophyta; Liliopsida; Poales; Poaceae;
            Ehrhartoideae; Oryzeae; Oryza.
REFERENCE   1  (bases 1 to 3168)
  AUTHORS   Ma,L., Wangc,J., Chen,C., Liu,X., Su,N., Li,L., Wang,X., Cao,M.,
            Jiao,Y., Sun,N., Zhang,X., Bao,J., Sun,D., Zhao,H., Yuan,L.,
            Wong,G.K.S., Deng,X.W. and Wang,J.
  TITLE     An analysis of transcriptional regulation of the rice genome and
            its comparison to Arabidopsis
  JOURNAL   Unpublished (2004)
COMMENT     Contact: Chen Chen
            Department of Bioinformatic
            Beijing Institute of Genomics
            Chinese Academy of Sciences, Beijing 101300, China
            Tel: 86-10-80481559
            Fax: 86-10-80488676
            Email: chenchen@genomics.org.cn
            Rice genomic sequence.
            Class: exon-trapped.
FEATURES             Location/Qualifiers
     source          1..3168
                     /organism="Oryza sativa (indica cultivar-group)"
                     /mol_type="genomic DNA"
                     /db_xref="taxon:39946"
                     /clone_lib="Oryza sativa Express Library"
                     /note="Oryza sativa exon trapped genomic sequences "
ORIGIN

  Query Match             17.9%;  Score 777;  DB 10;  Length 3168;
  Best Local Similarity   66.1%;  Pred. No. 4.9e-228;
  Matches 1178;  Conservative    0;  Mismatches  560;  Indels   45;  Gaps    2;

Qy     2217 GAGGATACTCTAAATATCGACATGTCAGACATTTATGACCATGATCCTGATCTATACGCA 2276

Db     1282 GGGGAGTCGCTCGACGTCAACGCCCACGACGTGTTCGACCACGACCCGGACCTCTACGGC 1341

Qy     2277 AAAATTGTTCGATACCCACTCGACATCATCCCCCTGTTGGACACTGAGTGTCAGGAAGTT 2336

Db     1342 AAGATGGTCAGGTACCCGCTCGAGGTGCTCGCCATCTTTGATATCGTGCTCATGGACCTC 1401

Qy     2337 GCTACCTCTTTACTACCAACGTTTGAGAAGCATATTGAGGCCAGACCTTTCAATCTCAAA 2396

Db     1402 GTCGCGCGCATCGAGCCCCTCTTCGAGAAGCACATCCAGACCAGGATCTACAACCTCAAG 1461

Qy     2397 GCATCGGTGCACATGCGTGAACTCAACCCTTCAGATATAGACAAATTGGTTTCTGTTAAA 2456

Db     1462 TCCTCGGTTTGCTTGAGGAATCTCAACCCTTCTGATATCGAGAAGATGGTGTCCATCAAG 1521

Qy     2457 GGAATGGTTATCCGGTGCAGTTCTATCATACCTGAAATTAAGGGGGCCTTCTTCAAATGT 2516

Db     1522 GGTATGATAATTCGGTGCAGCTCGGTGATTCCAGAGCTCAAGGAGGCTGTGTTCCGCTGC 1581

Qy     2517 TTAGTGTGTGGTCACTCGCCTCCGCTAGTTACAGTTGTTAAAGGGCGGGTTGAGGAGCCA 2576

Db     1582 TTGGTTTGCGGGTTCTACTCTGAGCCTGTAATGGTTGATCGAGGGAGGGTTACTGAGCCT 1641

Qy     2577 ACAAGGTGTGAAAAGCCAGAATGTGCAGCACGGAATGCTATGTCTCTTATTCACAATCGA 2636

Db     1642 CACATCTGTCAGAAAGAACAATGTAAAGCTACAAACTCTATGACTCTTGTGCATAACCGA 1701

Qy     2637 TGCACTTTTGCAAATAAGCAGATAGTGCGTCTTCAAGAAACTCCAGATGCCATTCCTGAA 2696

Db     1702 TGCAGGTTTGCAGATAAGCAGATCATAAAGTTGCAGGAAACACCAGATGAGATACCAGAA 1761

Qy     2697 GGAGAGACTCCACACACAGTCAGCATGTGTTTATACAACACTATGGTTGATGCTGTGAAG 2756

Db     1762 GGTGGCACTCCACATACTGTTAGTGTCTTGATGCACGATAAGCTCGTAGATGCTGGAAAG 1821

Qy     2757 CCTGGAGATCGTATTGAGGTAACAGGAGTTTTCAAGGCCATGGCAGTTCGAGTTGGTCCG 2816

Db     1822 CCTGGAGACAGGGTTGAGATAACTGGGATATACAGAGCCATGAGTATTAGAGTTGGCCCA 1881

Qy     2817 AATCAACGAACATTACGAGCATTGTATAAGACCTACATCGATTGCGTGCACGTCAAGAAG 2876

Db     1882 ACTCAGAGGACAGTGAAGTCGATTT-------------------------TCAAGAAG 1914

Qy     2877 TCTGACAGGGGTCGACTGCAAACTGAAGATCCTATGGAGATGGATAAGGAGAATGATATG 2936

Db     1915 ACAGACAAGTCTAGACTTCATGTTGAGGACTCCATGGAAACTGATAACCCCAATGCTA-- 1972

Qy     2937 TATGCTGGGTATCATGAAAGTGATACTTCAGAAGCTGCTAATGAAGCAAAGATTCAAAAA 2996

Db     1973 ----------------ACAAGACAACTGAAGATGATTTTCTCAGAGATAAGGTTGAGAAA 2016

Qy     2997 CTTAAAGAGCTGTCCAAGCTCCCGGGCATTTATGATAGACTTTCAAGGTCGCTGGCTCCA 3056

Db     2017 TTAAAAGAGTTGTCAAAGTTGCCAGATATATATGACAGATTAACTAGGTCATTGGCTCCA 2076

Qy     3057 AGCATTTGGGAGCTTGAAGATATTAAAAAGGGTCTTCTTTGCCAGCTCTTTGGTGGGAAG 3116

Db     2077 AACATATGGGAGCTGGACGACGTTAAAAGGGGCCTCCTTTGCCAGCTTTTTGGTGGAAAT 2136

Qy     3117 GCTAAGAAAATTCCATCTGGAGCATCTTTCCGAGGTGACATCAATGTTTTACTTGTTGGG 3176

Db     2137 GCTTTGAGGCTTCCTTCTGGAGCTAGTTTCCGAGGCGACATCAATATTTTGCTTGTTGGT 2196

Qy     3177 GACCCTGGTACCAGTAAATCTCAGCTGCTTCAGTATGTGCACAAGATAGCTCCTCGTGGA 3236

Db     2197 GATCCTGGAACAAGTAAATCCCAGCTTCTCCAATACATGCACAAACTGTCTCCTCGTGGC 2256

Qy     3237 ATCTACACTAGTGGGCGAGGAAGTTCGGCGGTTGGGCTGACAGCGTATGTAACGAAGGAT 3296

Db     2257 ATTTATACAAGTGGCAGAGGAAGTTCAGCTGTTGGCCTTACTGCATACGTTACCAAGGAT 2316

Qy     3297 CCAGAAACTCGAGAGACGGTATTGGAGAGCGGAGCTTTGGTTCTTAGTGATCGTGGGATA 3356

Db     2317 CCTGAAACTGGTGAAACTGTTCTTGAGAGTGGAGCGCTTGTTTTGAGTGACAAAGGTGTT 2376

Qy     3357 TGCTGTATCGATGAGTTCGACAAAATGTCTGATAATGCCCGAAGCATGCTTCATGAGGTA 3416

Db     2377 TGTTGTATTGATGAATTTGATAAGATGTCTGATAATGCTCGAAGCATGTTACATGAGGTG 2436

Qy     3417 ATGGAGCAACAAACGGTATCTGTAGCCAAAGGGGGTATCATTGCCTCGCTGAACGCTCGG 3476

Db     2437 ATGGAACAACAGACTGTCTCCATTGCCAAGGCTGGAATAATTGCATCTTTAAATGCCAGA 2496

Qy     3477 ACGTCTGTCCTTGCATGTGCAAATCCTAGTGGGTCCCGATACAATGCGCGCCTTTCTGTG 3536

Db     2497 ACATCAGTTCTAGCATGTGCAAATCCTACTGAATCACGTTATAATCCAAGGCTCTCTGTG 2556

Qy     3537 ATTGATAACATCCAGCTTCCTCCAACTCTACTTTCTAGATTTGATTTAATTTACTTAATG 3596

Db     2557 ATTGACAATATCCATCTTCCTCCAACACTGCTTTCTAGGTTTGACCTCATTTATCTGATA 2616

Qy     3597 CTCGACAAACCAGACGAGCAAAACGATCGTCGTCTCGCCAGGCATCTCGTGGCTTTACAC 3656

Db     2617 TTGGACAAGGCAGATGAGCAAACTGATAGACGCCTGGCTAAGCATATTGTTTCGTTGCAT 2676
```

```
Qy      3657 TATGAAAACTATGAAGTTTCAAAGCAGGACGCCTTAGATCTACAAACACTTACCGCGTAT 3716
Db      2677 TTTGAGAATCCAAACATAGAGGAGCTCGAGGTCTTGGATTTGCCAACACTTGTAGCCTAC 2736

Qy      3717 ATCACCTATGCTCGTCAGCATGTACATCCTACATTAAGTGATGAAGCTGCTGAAGATTTG 3776
Db      2737 ATAAGTTATGCAAGGAAGCATATACAACCACAGTTATCTGATGAAGCTGCAGAAGAATTG 2796

Qy      3777 ATTAATGGCTATGTTGAGATGCGCCAAAAGGGCAACTTTCCTGGAAGCAGTAAAAAGGTG 3836
Db      2797 ACCCGTGGCTATGTTGAGATGAGGAAAAGAGGAAACAGCCCTGGTAGCAGAAAGAAGGTC 2856

Qy      3837 ATAACAGCCACACCTCGGCAACTCGAAAGTATGATTCGTATCAGTGAAGCCCTAGCTCGA 3896
Db      2857 ATAACTGCGACAGCTCGACAAATTGAGAGCTTGATTCGGCTCAGTGAAGCACTGGCCCGA 2916

Qy      3897 ATGAGATTTTCTGAAGTGGTAGAGAAAGTTGATGCAGCAGAAGCTGTGCGCCTTTTAGAC 3956
Db      2917 ATGCGATTCTCTGAAGTGGTTGAAGTACGAGATGTTGTAGAGGCCTTCAGGCTTCTCGAA 2976

Qy      3957 GTCGCTTTGCAGCAATCTGCTACTGATCATGCAACAGGTACGA 3999
Db      2977 GTTGCCATGCAGCAATCGGCAACTGATCATGCCACTGCTTGGA 3019
```

```
RESULT 3
DR495517/c
LOCUS       DR495517                 879 bp    mRNA    linear   EST 06-JUL-2005
DEFINITION  WS0298.BR_E02 SS-IB-A-FL-15 Picea sitchensis cDNA clone WS0298_E02
            3, mRNA sequence.
ACCESSION   DR495517
VERSION     DR495517.1  GI:69453834
KEYWORDS    EST.
SOURCE      Picea sitchensis (Sitka spruce)
  ORGANISM  Picea sitchensis
            Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
            Spermatophyta; Coniferopsida; Coniferales; Pinaceae; Picea.
REFERENCE   1  (bases 1 to 879)
  AUTHORS   Ralph,S., Kolosova,N., Oddy,C., Cooper,D., Butterfield,Y.,
            Kirkpatrick,R., Liu,J., Palmquist,D., Stott,J., Barber,S., Yang,G.,
            Babakaiff,R., Brown-John,M., Chand,S., Featherstone,R., Masson,A.,
            Mayo,M., Moran,J., Olson,T., Wong,D., Friedmann,M.F., Ritland,C.E.,
            Siddiqui,A., Holt,R., Jones,S., Marra,M., Ellis,B.E., Douglas,C.,
            Ritland,K. and Bohlmann,J.
  TITLE     The spruce transcriptome: Analysis of expressed sequence tags from
            multiple cDNA libraries
  JOURNAL   Unpublished (2005)
COMMENT     Contact: Joerg Bohlmann
            Genome BC forest genomics program
            University of British Columbia
            Michael Smith Laboratories, 6174 University Boulevard, Rm. 237,
            Vancouver, British Columbia, Canada, V6T 1Z3
            Tel: 1-604-822-0282
            Fax: 1-604-822-2114
            Email: bohlmann@msl.ubc.ca
            Plate: WS0298  row: E  column: 02
            High quality sequence stop: 879.
FEATURES             Location/Qualifiers
     source          1..879
                     /organism="Picea sitchensis"
                     /mol_type="mRNA"
                     /cultivar="FB3-425"
                     /db_xref="taxon:3332"
                     /clone="WS0298_E02"
                     /sex="Hermaphrodite"
                     /lab_host="E. coli DH10B cells"
                     /clone_lib="SS-IB-A-FL-15"
                     /note="Organ: Bark (with phloem and cambium attached) from
                     two year old clonal trees grown under greenhouse
                     conditions in standard potting soil mixture; Vector:
                     pBluescript II SK (+) XR; Site_1: XhoI (5' end of cDNA);
                     Site_2: BamHI (3' end of cDNA); Adult white pine weevils
                     (Pissodes strobi) were caged with mesh bags on sapling
                     trees. Bark tissue with phloem and cambium attached was
                     harvested 2 hours, 6 hours and 48 hours after continuous
                     feeding by the insects. mRNA was isolated from each tissue
                     source independently and equal quantities of mRNA from
                     each tissue were then pooled. cDNA was prepared from 20
                     micrograms of mRNA according to the full-length cDNA
                     library construction method described by Carninci P. et
                     al. (2000), Genome Research 10(10):1617-1630 and
                     directionally ligated into the pBluescript II SK (+) XR
                     vector digested with XhoI (5' end) and BamHI (3'). Plasmid
                     DNA was then transformed by electroporation into DH10B
                     cells (Invitrogen) for propagation."
ORIGIN

  Query Match             11.0%;  Score 478;  DB 8;  Length 879;
  Best Local Similarity   71.5%;  Pred. No. 1.3e-135;
  Matches  628;  Conservative    0;  Mismatches  250;  Indels    0;  Gaps    0;

Qy      2415 GAACTCAACCCTTCAGATATAGACAAATTGGTTTCTGTTAAAGGAATGGTTATCCGGTGC 2474
Db       878 GAACTGAATCCATCAGATATTGAGAAAATGGTTTCAGTGAAAGGGATGATCATACGATGC 819

Qy      2475 AGTTCTATCATACCTGAAATTAAGGGGGCCTTCTTCAAATGTTTAGTGTGTGGTCACTCG 2534
Db       818 AGCTCAATAATACCTGAAATCAAAGAGGCATTTTTCAGATGCCTTGTTTGTGGCTATGCT 759

Qy      2535 CCTCCGCTAGTTACAGTTGTTAAAGGGCGGGTTGAGGAGCCAACAAGGTGTGAAAAGCCA 2594
Db       758 CCACCACCAGTGCCTGTTGATAGAGGTCGAATTGATGAACCTACTCGATGTGGAAGGCCA 699

Qy      2595 GAATGTGCAGCACGGAATGCTATGTCTCTTATTCACAATCGATGCACTTTTGCAAATAAG 2654
Db       698 GAATGTGCTGCTATGAACTCCATGACTTTAGTTCATAATAGATGCAGGTTTGCTGACAAA 639

Qy      2655 CAGATAGTGCGTCTTCAAGAAACTCCAGATGCCATTCCTGAAGGAGAGACTCCACACACA 2714
Db       638 CAGATCGTAAGGCTGCAAGAAACCCCTGATGCTATTCCAGAGGGAGAAACCCCACATACA 579

Qy      2715 GTCAGCATGTGTTTATACAACACTATGGTTGATGCTGTGAAGCCTGGAGATCGTATTGAG 2774
Db       578 GTCAGCTTGTTAATGCATGACAAACTAGTTGATGCTGCCAAGCCTGGTGACAGAGTTGAG 519

Qy      2775 GTAACAGGAGTTTTCAAGGCCATGGCAGTTCGAGTTGGTCCGAATCAACGAACATTACGA 2834
Db       518 GTAACAGGAATCTTCAGGGCCATGAGTGTTAGAGTAGGACCAAACCAGAGGACAGTCAAA 459

Qy      2835 GCATTGTATAAGACCTACATCGATTGCGTGCACGTCAAGAAGTCTGACAGGGGTCGACTG 2894
Db       458 TCCCTCTTTAAGACCTACATTGACTGCCTTCACTTGAAGAAGGCAGACAAGACACGACTG 399

Qy      2895 CAAACTGAAGATCCTATGGAGATGGATAAGGAGAATGATATGTATGCTGGGTATCATGAA 2954
Db       398 CAAGCTGAAGATCCAATGGAGCGTGATGCCACTAGTGACAGCTTGCTTCTTTTCAAGAG 339

Qy      2955 AGTGATACTTCAGAAGCTGCTAATGAAGCAAAGATTCAAAAACTTAAAGAGCTGTCCAAG 3014
Db       338 GGTGATATTTCACCCGTTCAGTATGAAGCCAAGATTGAGAAGTTGAAAGAGCTGTCACGG 279

Qy      3015 CTCCCGGGCATTTATGATAGACTTTCAAGGTCGCTGGCTCCAAGCATTTGGGAGCTTGAA 3074
Db       278 CAGCCTGATATATATGAGAAGTTAACCAGGTCATTGGCTCCAAGTATTTGGGAGCTGGAA 219

Qy      3075 GATATTAAAAAGGGTCTTCTTTGCCAGCTCTTTGGTGGGAAGGCTAAGAAAATTCCATCT 3134
Db       218 GATGTGAAAAAAGGGCTTTTATGTCAGCTTTTTGGGGAACTCCAGTAAAGCTAGCCACA 159

Qy      3135 GGAGCATCTTTCCGAGGTGACATCAATGTTTTACTTGTTGGGGACCCTGGTACCAGTAAA 3194
Db       158 GGTGCCACCTTCCGTGGTGACATCAACATTTTGCTTGTTGGGGACCCGGGTACGAGTAAG 99

Qy      3195 TCTCAGCTGCTTCAGTATGTGCACAAGATAGCTCCTCGTGGAATCTACACTAGTGGGCGA 3254
Db        98 TCTCAACTACTTCAGTATGTACACAAGTTGGCTCCTCGTGGTATTTATACAAGTGGTCGG 39
```

```
Qy       3255 GGAAGTTCGGCGGTTGGGCTGACAGCGTATGTAACGAA 3292

Db         38 GGAAGCTCTGCAGTTGGTTTGACTGCTTATGTGTCAAA 1


RESULT 4
CO123277
LOCUS       CO123277                 881 bp    mRNA    linear   EST 16-JUN-2004
DEFINITION  GR__Eb05F09.f GR__Eb Gossypium raimondii cDNA clone GR__Eb05F09 5',
            mRNA sequence.
ACCESSION   CO123277
VERSION     CO123277.1  GI:48821964
KEYWORDS    EST.
SOURCE      Gossypium raimondii
  ORGANISM  Gossypium raimondii
            Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
            Spermatophyta; Magnoliophyta; eudicotyledons; core eudicotyledons;
            rosids; eurosids II; Malvales; Malvaceae; Malvoideae; Gossypium.
REFERENCE   1  (bases 1 to 881)
  AUTHORS   Kim,H., Yu,Y., Kudrna,D., Hatfield,J., Stum,D., Mueller,C.,
            Udall,J.A., Rapp,R.A., Wendel,J.F., Rao,K., Soderlund,C. and
            Wing,R.A.
  TITLE     Global assembly of Cotton ESTs
  JOURNAL   Unpublished (2004)
COMMENT     Contact: Rod A. Wing
            Arizona Genomics Institute
            The University of Arizona
            Forbes Building Room 303, Tucson, AZ, 85721-0036, USA
            Tel: 520 626 9595
            Fax: 520 621 1259
            Email: rwing@genome.arizona.edu
            Plate: 05  row: F  column: 09.
FEATURES             Location/Qualifiers
     source          1..881
                     /organism="Gossypium raimondii"
                     /mol_type="mRNA"
                     /db_xref="taxon:29730"
                     /clone="GR__Eb05F09"
                     /tissue_type="floral"
                     /dev_stage="_3 to +3 DPA"
                     /lab_host="DH10B"
                     /clone_lib="GR__Eb"
                     /note="Vector: pCMV.SPORT-6.1; Site_1: NotI; Site_2:
                     EcoRV; Library made by Invitrogen with RNA supplied by
                     Wendle lab. Directional cloned into NotI-EV. Colonies
                     plated/picked by AGI. More glycerol clones held in -80."
ORIGIN

  Query Match             10.8%;  Score 469.6;  DB 7;  Length 881;
  Best Local Similarity   71.6%;  Pred. No. 5.3e-133;
  Matches  616;  Conservative    0;  Mismatches  244;  Indels    0;  Gaps    0;

Qy       3255 GGAAGTTCGGCGGTTGGGCTGACAGCGTATGTAACGAAGGATCCAGAAACTCGAGAGACG 3314

Db          1 GGGAGCTCTGCTGTTGGATTGACTGCTTATGTTGCTAAGGACCCTGAAACCGGAGAAACT 60

Qy       3315 GTATTGGAGAGCGGAGCTTTGGTTCTTAGTGATCGTGGGATATGCTGTATCGATGAGTTC 3374

Db         61 GTTTTGGAGAGTGGAGCCCTTGTTCTGAGTGACAGAGGCATCTGTTGTATTGATGAATTT 120

Qy       3375 GACAAAATGTCTGATAATGCCCGAAGCATGCTTCATGAGGTAATGGAGCAACAAACGGTA 3434

Db        121 GATAAAATGTCTGATAATGCAAGGAGCATGCTGCATGAGGTCATGGAACAACAGACTGTT 180

Qy       3435 TCTGTAGCCAAAGGGGGTATCATTGCCTCGCTGAACGCTCGGACGTCTGTCCTTGCATGT 3494

Db        181 TCAATAGCAAAAGCTGGTATAATTGCATCTCTTAATGCAAGGACTTCAGTCTTGGCCTGT 240

Qy       3495 GCAAATCCTAGTGGGTCCCGATACAATGCGCGCCTTTCTGTGATTGATAACATCCAGCTT 3554

Db        241 GCAAATCCTAGTGGTTCGCGCTATAATCCGCGTCTGTCTGTGATTGACAACATACACCTG 300

Qy       3555 CCTCCAACTCTACTTTCTAGATTTGATTTAATTTACTTAATGCTCGACAAACCAGACGAG 3614

Db        301 CCTCCAACCTTGCTGTCCAGATTTGATTTAATCTACTTGATTCTCGACAAGGCTGACGAG 360

Qy       3615 CAAAACGATCGTCGTCTCGCCAGGCATCTCGTGGCTTTACACTATGAAAACTATGAAGTT 3674

Db        361 CAGACAGATAGGCGTTTAGCTAAGCATATTGTTTCTTTACATTTTGAGGACACAGAGATT 420

Qy       3675 TCAAAGCAGGACGCCTTAGATCTACAAACACTTACCGCGTATATCACCTATGCTCGTCAG 3734

Db        421 GCTGAGCATGATGTCATAGATCTTGCTACATTAACAGCATATGTGAGCTATGCTCGCAAG 480

Qy       3735 CATGTACATCCTACATTAAGTGATGAAGCTGCTGAAGATTTGATTAATGGCTATGTTGAG 3794

Db        481 AATATTCACCCAAAATTATCTGATGAAGCTGCTGAAGAGCTAACCCGAGGATATGTTGAA 540

Qy       3795 ATGCGCCAAAAGGGCAACTTTCCTGGAAGCAGTAAAAAGGTGATAACAGCCACACCTCGG 3854

Db        541 ATGAGGAAAAGAGGAAATTTCCCAGGCAGCAGTAAAAAGGTCATCACAGCAACACCAAGG 600

Qy       3855 CAACTCGAAAGTATGATTCGTATCAGTGAAGCCCTAGCTCGAATGAGATTTTCTGAAGTG 3914

Db        601 CAAATTGAAAGTCTGATACGCCTTAGTGAAGCCTTGGCTCGTATTCGATTCTCAGAAACA 660

Qy       3915 GTAGAGAAAGTTGATGCAGCAGAAGCTGTGCGCCTTTTAGACGTCGCTTTGCAGCAATCT 3974

Db        661 GTTGAAATGCGAGATGTAACTGAGGCATTCCGGCTTTTGGAAGTTGCTATGCAGCAATCA 720

Qy       3975 GCTACTGATCATGCAACAGGTACGATAGACATGGATCTTATCACGACTGGAGTGTCGGCC 4034

Db        721 GCAACTGATCACTCCACAGGAACTATTGATATGGATCTTATCACCACTGGAGTTTCTGCA 780

Qy       4035 AGCGAGCGTATTCGTCGGGCCAACTTGCTAGCTGCTCTGCGAGAGCTTATAGCAGATAAA 4094

Db        781 AGTGAAAGGATGAGGAGGGAGAGTGTGTTATCAGCAACACGCAACGTAATAATGGATAAG 840

Qy       4095 ATTTCACCTGGCAGCTCCTC 4114

Db        841 CTACAACTTGGGGGACCATC 860


RESULT 5
CK210237
LOCUS       CK210237                1156 bp    mRNA    linear   EST 08-DEC-2003
DEFINITION  FGAS022036 Triticum aestivum FGAS: Library 5 GATE 7 Triticum
            aestivum cDNA, mRNA sequence.
ACCESSION   CK210237
VERSION     CK210237.1  GI:39572627
KEYWORDS    EST.
SOURCE      Triticum aestivum (bread wheat)
  ORGANISM  Triticum aestivum
            Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
            Spermatophyta; Magnoliophyta; Liliopsida; Poales; Poaceae;
            Pooideae; Triticeae; Triticum.
REFERENCE   1  (bases 1 to 1156)
  AUTHORS   Allard,F., Crosby,W.L., Danyluk,J., Eudes,F., Frick,M., Gaudet,D.,
            Genswein,B., Graf,R., Gulick,P., Hrycan,L.D., Laroche,A.,
            Links,M.G., McCarthy,E.L., Monroy,A., Muzak,I., Nilson,D.,
            Penniket,C., Roach,J.L. and Sarhan,F.
  TITLE     Functional Genomics of Abiotic Stress In Wheat and Canola Crops
  JOURNAL   Unpublished (2003)
COMMENT     Contact: Wm L Crosby
            Bioinformatics
            University of Saskatchewan, Department of Computer Science
            1C101 Engineering Building, 57 Campus Drive, Saskatoon,
            Saskatchewan, S7N 5A9, Canada
            Tel: 306 966 1769
            Fax: 306 966 2033
            Email: fgas_ests@cs.usask.ca
            This sequence is the direct result of the Base calling software
            Phred (default parameters). It is the raw base calls. To aid in the
            identification of the high quality insert the software Lucy
            (default parameters) has been run on this sequence. Lucy identified
```

```
            the region [16,727].
            Plate: L5B022  row: D  column: 10.
FEATURES             Location/Qualifiers
     source          1..1156
                     /organism="Triticum aestivum"
                     /mol_type="mRNA"
                     /db_xref="taxon:4565"
                     /clone_lib="Triticum aestivum FGAS: Library 5 GATE 7"
                     /note="Vector: pCMV.SPORT6; Crown and developmental stages
                     of spike formation in wheat cultivar Norstar. 4 mRNA
                     populations were combined before constructing the library.
                     The first mRNA population is from 1cm crown sections after
                     30 days of cold acclimation. The second is from 1cm crown
                     sections after 11 days of deacclimation (before
                     deacclimation plants were fully vernalized for 49 days).
                     The third is from different developmental stages of spike
                     formation (5 to 50mm) that still have not emerged from the
                     leaf (dissection required). The last is from different
                     developmental stages of spike and seed formation after
                     having emerged from the leaf (visible). First strand
                     synthesis in this library was done in the presence of
                     methylated dCTP thereby protecting from internal cleavage
                     with NotI."
ORIGIN

  Query Match             10.5%;  Score 456.4;  DB 7;  Length 1156;
  Best Local Similarity   68.5%;  Pred. No. 7.2e-129;
  Matches  659;  Conservative    0;  Mismatches  301;  Indels    2;  Gaps    2;

Qy       3160 ATGTTTTACTTGTTGGGGACCCTGGTACCAGTAAATCTCAGCTGCTTCAGTATGTGCACA 3219
Db         36 ATATTTTGCTTGTTGGTGATCCTGGAACGAGCAAATCCCATCTTCTCCAGTACATGCATA 95

Qy       3220 AGATAGCTCCTCGTGGAATCTACACTAGTGGGCGAGGAAGTTCGGCGGTTGGGCTGACAG 3279
Db         96 AACTGTCTCCTCGTGGTATTTACACAAGTGGAAGAGGCAGTTCAGCAGTTGGCCTTACTG 155

Qy       3280 CGTATGTAACGAAGGATCCAGAAACTCGAGAGACGGTATTGGAGAGCGGAGCTTTGGTTC 3339
Db        156 CTTATGTTGCTAAGGACCCTGAAACTGGTGAAACTGTTCTTGAAAGTGGAGCACTTGTTT 215

Qy       3340 TTAGTGATCGTGGGATATGCTGTATCGATGAGTTCGACAAAATGTCTGATAATGCCCGAA 3399
Db        216 TGAGTGACAAAGGTGTTTGCTGTATTGATGAGTTTGATAAGATGTCTGATAATGCCCGAA 275

Qy       3400 GCATGCTTCATGAGGTAATGGAGCAACAAACGGTATCTGTAGCCAAAGGGGGTATCATTG 3459
Db        276 GCATGCTGCATGAGGTGATGGAGCAGCAGACTGTATCCATTGCAAAGGCTGGAATTATTG 335

Qy       3460 CCTCGCTGAACGCTCGGACGTCTGTCCTTGCATGTGCAAATCCTAGTGGGTCCCGATACA 3519
Db        336 CATCTTTGAATGCTAGGACATCTGTACTAGCATGTGCCAATCCATCTGAATCACGTTACA 395

Qy       3520 ATGCGCGCCTTTCTGTGATTGATAACATCCAGCTTCCTCCAACTCTACTTTCTAGATTTG 3579
Db        396 ATCCAAGGCTTTCTGTGATCGACAATATCCACCTTCCTCCAACACTGCTGTCAAGGTTTG 455

Qy       3580 ATTTAATTTACTTAATGCTCGACAAACCAGACGAGCAAAACGATCGTCGTCTCGCCAGGC 3639
Db        456 ACCTGATTTATCTGATCTTGGACAAGGCAGACGAACAAACTGATAGACGCCTGGCTAAGC 515

Qy       3640 ATCTCGTGGCTTTACACTATGAAAACTATGAAGTTTCAAAGCAGGACGCCTTAGATCTAC 3699
Db        516 ATATTGTTTCATTGCATTTCGAGAATCCAGAAGTAGTTGAGCACCAGGTCTTGGATTTGC 575

Qy       3700 AAACACTTACCGCGTATATCACCTATGCTCGTCAGCATGTACATCCTACATTAAGTGATG 3759
Db        576 CCACGTTAGTTGCGTACATCAGCTATGCAAGGAAGTACATTCAGCCAAAGTTATCTGATG 635

Qy       3760 AAGCTGCTGAAGATTTGATTAATGGCTATGTTGAGATGCGCCAAAAGGGCAACTTTCCTG 3819
Db        636 AAGCTGCAGAAGAATTGACCCGTGGCTACGTTGCAATGAGGCAAAGGGGGAACAATCCTG 695

Qy       3820 GAAGCAGTAAAAAGGTGATAACAGCCACACCTCGGCAACTCGAAAGTATGATTCGTATCA 3879
Db        696 GTAGCAGAAAGAAGGTCATCACAGCAACAGCTAGGCCAATTGAGAGCTTGATTCGTCTTA 755

Qy       3880 GTGAAGCCCTAGCTCGAATGAGATTTTCTGAAGTGGTAGAGAAAGTTGATGCAGCAGAAG 3939
Db        756 GTGAAGCACTGGCGCAAATGCTTTTTCAGAAGTTGGTGGAGTGCGAGATGTAATATCAAG 815

Qy       3940 CTGTGCGCCTTTTAGACGTCGCTTTGCAGCAATCTGCTACTGATCATGCAACAGGTACG- 3998
Db        816 CTTTAGGCTTCTTGGAAGTCGCCATGCAGCAATCTTGAACCGATCATGCAACAGGGACGA 875

Qy       3999 ATAGACATGGATCTTATCACGACTGGAGTGTCGGCCAGCGAGCGTATTCGTCGGGCCAAC 4058
Db        876 ATTGATATGGATTCTATCATGACTGGGGTATC-CCAAGGTAAAAGGCGAAACGTGAAATC 934

Qy       4059 TTGCTAGCTGCTCTGCGAGAGCTTATAGCAGATAAAATTTCACCTGGCAGCTCCTCTGGC 4118
Db        935 TTGTTGGGGGCTATCCGAAACTTGTTATGGAAAAAAAGCCAGTTTGGGGGCCCCTGATGC 994

Qy       4119 TT 4120
Db        995 GT 996


RESULT 6
DR503163
LOCUS       DR503163                 907 bp    mRNA    linear   EST 07-JUL-2005
DEFINITION  WS0298.B21_E02 SS-IB-A-FL-15 Picea sitchensis cDNA clone WS0298_E02
            5, mRNA sequence.
ACCESSION   DR503163
VERSION     DR503163.1  GI:70241689
KEYWORDS    EST.
SOURCE      Picea sitchensis (Sitka spruce)
  ORGANISM  Picea sitchensis
            Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
            Spermatophyta; Coniferopsida; Coniferales; Pinaceae; Picea.
REFERENCE   1  (bases 1 to 907)
  AUTHORS   Ralph,S., Kolosova,N., Oddy,C., Cooper,D., Butterfield,Y.,
            Kirkpatrick,R., Liu,J., Palmquist,D., Stott,J., Barber,S., Yang,G.,
            Babakaiff,R., Brown-John,M., Chand,S., Featherstone,R., Masson,A.,
            Mayo,M., Moran,J., Olson,T., Wong,D., Friedmann,M.F., Ritland,C.E.,
            Siddiqui,A., Holt,R., Jones,S., Marra,M., Ellis,B.E., Douglas,C.,
            Ritland,K. and Bohlmann,J.
  TITLE     The spruce transcriptome: Analysis of expressed sequence tags from
            multiple cDNA libraries
  JOURNAL   Unpublished (2005)
COMMENT     Contact: Joerg Bohlmann
            Genome BC forest genomics program
            University of British Columbia
            Michael Smith Laboratories, 6174 University Boulevard, Rm. 237,
            Vancouver, British Columbia, Canada, V6T 1Z3
            Tel: 1-604-822-0282
            Fax: 1-604-822-2114
            Email: bohlmann@msl.ubc.ca
            Plate: WS0298  row: E  column: 02
            High quality sequence stop: 907.
FEATURES             Location/Qualifiers
     source          1..907
                     /organism="Picea sitchensis"
                     /mol_type="mRNA"
                     /cultivar="FB3-425"
                     /db_xref="taxon:3332"
                     /clone="WS0298_E02"
                     /sex="Hermaphrodite"
                     /lab_host="E. coli DH10B cells"
                     /clone_lib="SS-IB-A-FL-15"
                     /note="Organ: Bark (with phloem and cambium attached) from
                     two year old clonal trees grown under greenhouse
                     conditions in standard potting soil mixture; Vector:
                     pBluescript II SK (+) XR; Site_1: XhoI (5' end of cDNA);
                     Site_2: BamHI (3' end of cDNA); Adult white pine weevils
                     (Pissodes strobi) were caged with mesh bags on sapling
```

```
            trees. Bark tissue with phloem and cambium attached was
            harvested 2 hours, 6 hours and 48 hours after continuous
            feeding by the insects. mRNA was isolated from each tissue
            source independently and equal quantities of mRNA from
            each tissue were then pooled. cDNA was prepared from 20
            micrograms of mRNA according to the full-length cDNA
            library construction method described by Carninci P. et
            al. (2000), Genome Research 10(10):1617-1630 and
            directionally ligated into the pBluescript II SK (+) XR
            vector digested with XhoI (5' end) and BamHI (3'). Plasmid
            DNA was then transformed by electroporation into DH10B
            cells (Invitrogen) for propagation."
ORIGIN

  Query Match             10.1%;  Score 439.4;  DB 8;  Length 907;
  Best Local Similarity   67.8%;  Pred. No. 1.2e-123;
  Matches  614;  Conservative    0;  Mismatches  291;  Indels    0;  Gaps    0;

Qy       2178 AAGTACATCCAGATCATAGAGGAGACTGTGGAGCGTGAGGAGGATACTCTAAATATCGAC 2237
Db          3 AAATACATGAGGATTTTAGAGCAGACAATGGAAGTAGAAGGCGATTCTTTGAACGTTGAT 62

Qy       2238 ATGTCAGACATTTATGACCATGATCCTGATCTATACGCAAAAATTGTTCGATACCCACTC 2297
Db         63 GCAATGAACATCTTGGATTATGATGAGGACCTTTATGGGAAAATGGTTAGATACCCCCTC 122

Qy       2298 GACATCATCCCCCTGTTGGACACTGAGTGTCAGGAAGTTGCTACCTCTTTACTACCAACG 2357
Db        123 GAAGTTCTTGCCATTTTTGATATTGTTTTGATGGATGTTGCTGTTCGCCTTAAACCTTTG 182

Qy       2358 TTTGAGAAGCATATTGAGGCCAGACCTTTCAATCTCAAAGCATCGGTGCACATGCGTGAA 2417
Db        183 TGGGAGAAGCACATACAGGCAAGAATATTCAATCTCAAAGATTCAGTTAACATGCGAGAA 242

Qy       2418 CTCAACCCTTCAGATATAGACAAATTGGTTTCTGTTAAAGGAATGGTTATCCGGTGCAGT 2477
Db        243 CTGAATCCATCAGATATTGAGAAAATGGTTTCAGTGAAAGGGATGATCATACGATGCAGC 302

Qy       2478 TCTATCATACCTGAAATTAAGGGGGCCTTCTTCAAATGTTTAGTGTGTGGTCACTCGCCT 2537
Db        303 TCAATAATACCTGAAATCAAAGAGGCATTTTTCAGATGCCTTGTTTGTGGCTATGCTCCA 362

Qy       2538 CCGCTAGTTACAGTTGTTAAAGGGCGGGTTGAGGAGCCAACAAGGTGTGAAAAGCCAGAA 2597
Db        363 CCACCAGTGCCTGTTGATAGAGGTCGAATTGATGAACCTACTCGATGTGGAAGGCCAGAA 422

Qy       2598 TGTGCAGCACGGAATGCTATGTCTCTTATTCACAATCGATGCACTTTTGCAAATAAGCAG 2657
Db        423 TGTGCTGCTATGAACTCCATGACTTTAGTTCATAATAGATGCAGGTTTGCTGACAAACAG 482

Qy       2658 ATAGTGCGTCTTCAAGAAACTCCAGATGCCATTCCTGAAGGAGAGACTCCACACACAGTC 2717
Db        483 ATCGTAAGGCTGCAAGAAACCCCTGATGCTATTCCAGAGGGAGAAACCCCACATACAGTC 542

Qy       2718 AGCATGTGTTTATACAACACTATGGTTGATGCTGTGAAGCCTGGAGATCGTATTGAGGTA 2777
Db        543 AGCTTGTTAATGCATGACAAACTAGTTGATGCTGCCAAGCCTGGTGACAGAGTTGAGGTA 602

Qy       2778 ACAGGAGTTTTCAAGGCCATGGCAGTTCGAGTTGGTCCGAATCAACGAACATTACGAGCA 2837
Db        603 ACAGGAATCTTCAGGGCCATGAGTGTTAGAGTAGGACCAAACCAGAGGACAGTCAAATCC 662

Qy       2838 TTGTATAAGACCTACATCGATTGCGTGCACGTCAAGAAGTCTGACAGGGGTCGACTGCAA 2897
Db        663 CTCTTTAAGACCTACATTGACTGCCTTCACTTGAAGAAGGCAGACAAGACACGACTGCAA 722

Qy       2898 ACTGAAGATCCTATGGAGATGGATAAGGAGAATGATATGTATGCTGGGTATCATGAAAGT 2957
Db        723 GCTGAAGATCCAATGGAGCGTGATGCCACTAGTGACAGCTTTGCTTCTTTTCAAGAGGGT 782

Qy       2958 GATACTTCAGAAGCTGCTAATGAAGCAAAGATTCAAAAACTTAAAGAGCTGTCCAAGCTC 3017
Db        783 GATATTTCACCCGTTCAGTATGAAGCCAAGATTGAGAAGTTGAAAGAGCTGTCACGGCAG 842

Qy       3018 CCGGGCATTTATGATAGACTTTCAAGGTCGCTGGCTCCAAGCATTTGGGAGCTTGAAGAT 3077
Db        843 CCTGATATATATGAGAAGTTAAACCAGTCATTGGCTCCAAGTATTTGGGAGCTGGAAGAT 902

Qy       3078 ATTAA 3082
Db        903 GTGAA 907


RESULT 7
BF051102
LOCUS       BF051102                 770 bp    mRNA    linear   EST 18-MAY-2001
DEFINITION  EST436277 tomato developing/immature green fruit Lycopersicon
            esculentum cDNA clone cLEM21H9 5' sequence, mRNA sequence.
ACCESSION   BF051102
VERSION     BF051102.1  GI:10804998
KEYWORDS    EST.
SOURCE      Lycopersicon esculentum (Solanum lycopersicum)
  ORGANISM  Lycopersicon esculentum
            Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
            Spermatophyta; Magnoliophyta; eudicotyledons; core eudicotyledons;
            asterids; lamiids; Solanales; Solanaceae; Solanum; Lycopersicon.
REFERENCE   1  (bases 1 to 770)
  AUTHORS   Alcala,J., Vrebalov,J., White,R., van der Hoeven,R.S., Holt,I.E.,
            Liang,F., Hansen,T.S., Craven,M.B., Bowman,C.L., Ronning,C.M.,
            Nierman,W., Fraser,C.M., Martin,G.B., Giovannoni,J.J. and
            Tanksley,S.D.
  TITLE     Generation of ESTs from tomato fruit tissue, immature green
  JOURNAL   Unpublished (2000)
COMMENT     Contact: CUGI
            Clemson University Genomics Institute
            Clemson University
            100 Jordan Hall, Clemson, SC 29634, USA
            Email: http://www.genome.clemson.edu/orders/index.html.
FEATURES             Location/Qualifiers
     source          1..770
                     /organism="Lycopersicon esculentum"
                     /mol_type="mRNA"
                     /cultivar="TA496"
                     /db_xref="taxon:4081"
                     /clone="cLEM21H9"
                     /tissue_type="fruit"
                     /dev_stage="immature green (5-35 days post-anthesis)"
                     /lab_host="SOLR"
                     /clone_lib="tomato developing/immature green fruit"
                     /note="Vector: pBluescriptSKmCUadapt; Site_1: EcoR1;
                     Site_2: Xho1; Fruit were tagged at 5 dpa (0.5 cm) and
                     harvested at 7 day intervals through 35 dpa.  Equal masses
                     of tissue from each stage were combined (including seeds
                     and locules) prior to mRNA isolation."
ORIGIN

  Query Match             10.0%;  Score 434.2;  DB 2;  Length 770;
  Best Local Similarity   73.5%;  Pred. No. 4.6e-122;
  Matches  567;  Conservative    0;  Mismatches  203;  Indels    1;  Gaps    1;

Qy       2996 ACTTAAAGAGCTGTCCAAGCTCCCGGGCATTTATGATAGACTTTCAAGGTCGCTGGCTCC 3055
Db          1 ATTGATAGAACTATCTAAACAGCCTGATATCTATGAAAGGCTGACCAGGTCCTTGGCACC 60

Qy       3056 AAGCATTTGGGAGCTTGAAGATATTAAAAAGGGTCTTCTTTGCCAGCTCTTTGGTGGGAA 3115
Db         61 AAACATATGGGAGTTGGATGATGTTAAGAAAGGCCTTCTTTGCCAGCTCTTTGGTGGAAA 120

Qy       3116 GGCTAAGAAAATTCCATCTGGAGCATCTTTCCGAGGTGACATCAATGTTTTACTTGTTGG 3175
Db        121 TGCATTGACCTTGCCGTCTGGAGCAAGCTTCCGTGGTGATATCAACATTCTTCTTGTTGG 180

Qy       3176 GGACCCTGGTACCAGTAAATCTCAGCTGCTTCAGTATGTGCACAAGATAGCTCCTCGTGG 3235
Db        181 TGATCCGGGGACCAGCAAATCACAACTGCTCCAGTATATTCACAAACTGTCTCCTCGTGG 240
```

```
Qy       3236 AATCTACACTAGTGGGCGAGGAAGTTCGGCGGTTGGGCTGACAGCGTATGTAACGAAGGA 3295
Db        241 TATATACACAAGTGGACGAGGAAGTTCAGCTGTGGGGTTGACTGCTTATGTAGCCAAAGA 300

Qy       3296 TCCAGAAACTCGAGAGACGGTATTGGAGAGCGGAGCTTTGGTTCTTAGTGATCGTGGGAT 3355
Db        301 TCCCGAGACTGGTGAAACTGTCCTTGAGAGTGGCGCTCTGGTCTTGAGTGACAGAGGGAT 360

Qy       3356 ATGCTGTATCGATGAGTTCGACAAAATGTCTGATAATGCCCGAAGCATGCTTCATGAGGT 3415
Db        361 CTGCTGTATTGACGAGTTTGACAAAATGTCTGACAGTGCCAGGAGCATGTTACATGAGGT 420

Qy       3416 AATGGAGCAACAAACGGTATCTGTAGCCAAAGGGGGTATCATTGCCTCGCTGAACGCTCG 3475
Db        421 GATGGAGCAACAAACTGTTTCAATTGCAAAGGCTGGAATTATCGCTTCGCTTAATGCTAG 480

Qy       3476 GACGTCTGTCCTTGCATGTGCAAATCCTAGTGGGTCCCGATACAATGCGCGCCTTTCTGT 3535
Db        481 GACTTCAGTATTGGCATGTGCAAATCCAATTGGCTCCCGTTACAATCCCCGATTATCTGT 540

Qy       3536 GATTGATAACATCCAGCTTCCTCCAACTCTACTTTCTAGATTTGATTTAATTTACTTAAT 3595
Db        541 CATTGATAATATACACCTTCCACCTACCCTGCTGTCTAGATTTGATTTGATATATTTAAT 600

Qy       3596 GCTCGACAAACCAGACGAGCAAAACGATCGTCGTCTCGCCAGGCATCTCGTGGCTTTACA 3655
Db        601 TCTAGACAAAGCAGATGAGCAGACCGACAGGCGCCTTGCAAAGCACATAGTTGCTTTACA 660

Qy       3656 CTATGAAAACTATGAAGTTTCAAAGCAGGACGCCTTAGATCTACAAACACTTACCGCGTA 3715
Db        661 CTCTGAGAATCCTGAGAACTCTGAGCAAGAAGTGATTGACCTTCCAACATTAGCTTCATA 720

Qy       3716 TATCACCTATGCTCGTCAGCATGTACATCCTACATTAAGTGATGAAGCTGC 3766
Db        721 CTTGAGCTATGCTCG-AAACATATACATCCACAATTATCTGATGAAGCAGC 770
```

```
RESULT 8
BM412106
LOCUS       BM412106                788 bp    mRNA    linear   EST 22-JAN-2002
DEFINITION  EST586433 tomato breaker fruit Lycopersicon esculentum cDNA clone
            cLEG59A13 5' end, mRNA sequence.
ACCESSION   BM412106
VERSION     BM412106.1  GI:18263736
KEYWORDS    EST.
SOURCE      Lycopersicon esculentum (Solanum lycopersicum)
  ORGANISM  Lycopersicon esculentum
            Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
            Spermatophyta; Magnoliophyta; eudicotyledons; core eudicotyledons;
            asterids; lamiids; Solanales; Solanaceae; Solanum; Lycopersicon.
REFERENCE   1  (bases 1 to 788)
  AUTHORS   Alcala,J., Vrebalov,J., White,R., Vision,T., Karamycheva,S.A.,
            Tsai,J., Bougri,O., Kirkness,E., Utterback,T., Van Aken,S.,
            Ronning,C.M., Fraser,C.M., Martin,G.B., Tanksley,S.D. and
            Giovannoni,J.
  TITLE     Generation of ESTs from tomato fruit tissue, breaker stage (2002)
  JOURNAL   Unpublished (2002)
COMMENT     Contact: CUGI
            Clemson University Genomics Institute
            Clemson University
            100 Jordan Hall, Clemson, SC 29634, USA
            Email: http://www.genome.clemson.edu/orders/index.html
            This clone is available through the Clemson University Genomics
            Institute
            Seq primer: T3.
FEATURES             Location/Qualifiers
     source          1..788
                     /organism="Lycopersicon esculentum"
                     /mol_type="mRNA"
                     /cultivar="TA496"
                     /db_xref="taxon:4081"
                     /clone="cLEG59A13"
                     /tissue_type="Pericarp"
                     /dev_stage="breaker"
                     /lab_host="SOLR"
                     /clone_lib="tomato breaker fruit"
                     /note="Vector: pBluescriptSKmCUadapt; Site_1: EcoR1;
                     Site_2: XhoI; supplier: Boyce Thompson Institute;
                     sequencing: The Institute for Genomic Research.  Fruit
                     were harvested at the breaker stage (first sign of
                     lycopene accumulation on the blossom end of fruit).  Fruit
                     were cut in half and the seeds and locules were discarded
                     prior to freezing the pericarp."
ORIGIN

  Query Match             10.0%;  Score 433;  DB 3;  Length 788;
  Best Local Similarity   72.0%;  Pred. No. 1.1e-121;
  Matches  565;  Conservative    0;  Mismatches  220;  Indels    0;  Gaps    0;
```

```
Qy       3321 GAGAGCGGAGCTTTGGTTCTTAGTGATCGTGGGATATGCTGTATCGATGAGTTCGACAAA 3380
Db          2 GAGAGTGGCGCTCTGGTCTTGAGTGACAGAGGGATCTGCTGTATTGACGAGTTTGACAAA 61

Qy       3381 ATGTCTGATAATGCCCGAAGCATGCTTCATGAGGTAATGGAGCAACAAACGGTATCTGTA 3440
Db         62 ATGTCTGACAGTGCCAGGAGCATGTTACATGAGGTGATGGAGCAACAAACTGTTTCAATT 121

Qy       3441 GCCAAAGGGGGTATCATTGCCTCGCTGAACGCTCGGACGTCTGTCCTTGCATGTGCAAAT 3500
Db        122 GCAAAGGCTGGAATTATCGCTTCGCTTAATGCTAGGACTTCAGTATTGGCATGTGCAAAT 181

Qy       3501 CCTAGTGGGTCCCGATACAATGCGCGCCTTTCTGTGATTGATAACATCCAGCTTCCTCCA 3560
Db        182 CCAATTGGCTCCCGTTACAATCCCCGATTATCTGTCATTGATAATATACACCTTCCACCT 241

Qy       3561 ACTCTACTTTCTAGATTTGATTTAATTTACTTAATGCTCGACAAACCAGACGAGCAAAAC 3620
Db        242 ACCCTGCTGTCTAGATTTGATTTGATATATTTAATTCTAGACAAAGCAGATGAGCAGACC 301

Qy       3621 GATCGTCGTCTCGCCAGGCATCTCGTGGCTTTACACTATGAAAACTATGAAGTTTCAAAG 3680
Db        302 GACAGGCGCCTTGCAAAGCACATAGTTGCTTTACACTCTGAGAATCCTGAGAACTCTGAG 361

Qy       3681 CAGGACGCCTTAGATCTACAAACACTTACCGCGTATATCACCTATGCTCGTCAGCATGTA 3740
Db        362 CAAGAAGTGATTGACCTTCCAACATTAGCTTCATACTTGAGCTATGCTCGAAAACATATA 421

Qy       3741 CATCCTACATTAAGTGATGAAGCTGCTGAAGATTTGATTAATGGCTATGTTGAGATGCGC 3800
Db        422 CATCCACAATTATCTGATGAAGCAGCTGAAGAATTGACTAGAGGTTATGTGGAGATGAGA 481

Qy       3801 CAAAAGGGCAACTTTCCTGGAAGCAGTAAAAAGGTGATAACAGCCACACCTCGGCAACTC 3860
Db        482 AGAAAAGGGAATTTCCCCGGTAGCAGTAAAAAGGTGATTACAGCTACACCAAGGCAGCTA 541

Qy       3861 GAAAGTATGATTCGTATCAGTGAAGCCCTAGCTCGAATGAGATTTTCTGAAGTGGTAGAG 3920
Db        542 GAGAGTTTGATACGTCTAGGTGAAGGTCTTGCTCGGATGCGTTTTTCAGAAAAGGTGGAG 601

Qy       3921 AAAGTTGATGCAGCAGAAGCTGTGCGCCTTTTAGACGTCGCTTTGCAGCAATCTGCTACT 3980
Db        602 AAAAGAGATGTAGTGGAGGCTTTTCGACTTCTAGAAGTTGCATTGCAGCAGTCTGCAACT 661

Qy       3981 GATCATGCAACAGGTACGATAGACATGGATCTTATCACGACTGGAGTGTCGGCCAGCGAG 4040
Db        662 GACCATGCTACAGGAACCATTGACATGGATCTCATCACAACTGGAGTATCTGCAAGTGAA 721

Qy       4041 CGTATTCGTCGGGCCAACTTGCTAGCTGCTCTGCGAGAGCTTATAGCAGATAAAATTTCA 4100
Db        722 AGGATGAGAAGGGAGAATTTGGTGTCAAGCACCCGCAACATAATTATGGAGAAGATGCAG 781

Qy       4101 CCTGG 4105
Db        782 CTTGG 786
```

```
RESULT 9
AY408033
LOCUS       AY408033                2522 bp    DNA     linear   GSS 12-DEC-2003
DEFINITION  Homo sapiens HCM3089 gene, VIRTUAL TRANSCRIPT, partial sequence,
            genomic survey sequence.
ACCESSION   AY408033
VERSION     AY408033.1  GI:39764004
KEYWORDS    GSS.
SOURCE      Homo sapiens (human)
  ORGANISM  Homo sapiens
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Euarchontoglires; Primates; Catarrhini;
            Hominidae; Homo.
REFERENCE   1  (bases 1 to 2522)
  AUTHORS   Clark,A.G., Glanowski,S., Nielson,R., Thomas,P., Kejariwal,A.,
            Todd,M.A., Tanenbaum,D.M., Civello,D.R., Lu,F., Murphy,B.,
            Ferriera,S., Wang,G., Zheng,X.H., White,T.J., Sninsky,J.J.,
            Adams,M.D. and Cargill,M.
  TITLE     Inferring nonneutral evolution from human-chimp-mouse orthologous
            gene trios
  JOURNAL   Science 302 (5652), 1960-1963 (2003)
   PUBMED   14671302
REFERENCE   2  (bases 1 to 2522)
  AUTHORS   Clark,A.G., Glanowski,S., Nielson,R., Thomas,P., Kejariwal,A.,
            Todd,M.A., Tanenbaum,D.M., Civello,D.R., Lu,F., Murphy,B.,
            Ferriera,S., Wang,G., Zheng,X.H., White,T.J., Sninsky,J.J.,
            Adams,M.D. and Cargill,M.
  TITLE     Direct Submission
  JOURNAL   Submitted (16-NOV-2003) Celera Genomics, 45 West Gude Drive,
            Rockville, MD 20850, USA
COMMENT     This sequence was made by sequencing genomic exons and ordering
            them based on alignment.
FEATURES             Location/Qualifiers
     source          1..2522
                     /organism="Homo sapiens"
                     /mol_type="genomic DNA"
                     /db_xref="taxon:9606"
     gene            <1..>2522
                     /locus_tag="HCM3089"
ORIGIN

  Query Match             9.9%;  Score 429.2;  DB 10;  Length 2522;
  Best Local Similarity   53.6%;  Pred. No. 2.6e-120;
  Matches 1113;  Conservative    0;  Mismatches  878;  Indels   87;  Gaps    7;

Qy       2181 TACATCCAGATCATAGAGGAGACTGTGGAGCGTGAGGAGGATACTCTAAATATCGACATG 2240
Db        507 TACATGCAACGACTTGGGGAGATTAATGTTATTGGTGAGCCATTTTTAAATGTGAACTGT 566
Qy       2241 TCAGACATTTATGACCATGATCCTGATCTATACGCAAAAATTGTTCGATACCCACTCGAC 2300
Db        567 GAACACATCAAATCATTTGACAAAAATTTGTACAGACAACTCATCTCTTACCCACAGGAA 626
Qy       2301 ATCATCCCCCTGTTGGACACTGAGTGTCAGGAA------GTTGCTACCTCTTTACTACCA 2354
Db        627 GTTATTCCAACTTTTGACATGGCTGTCAATGAAATCTTCTTTGACCGTTACCCTGACTCA 686
Qy       2355 ACGTTTGAGAAGCATATTGAGGCCAGACCTTTCAATCTCAAAGCATCGGTGCACATGCGT 2414
Db        687 ATCTTAGAACATCAGATTCAAGTAAGACCATTCAACGCATTGAAGACTAAGAATATGAGA 746
Qy       2415 GAACTCAACCCTTCAGATATAGACAAATTGGTTTCTGTTAAAGGAATGGTTATCCGGTGC 2474
Db        747 AACCTGAATCCAGAAGACATTGACCAGCTCATCACCATCAGCGGCATGGTGATCAGGACA 806
Qy       2475 AGTTCTATCATACCTGAAATTAAGGGGGCCTTCTTCAAATGTTTAGTGTGTGGTCACTCG 2534
Db        807 TCCCAGCTGATTCCCGAGATGCAGGAGGCCTTCTTCCAGTGCCAAGTGTGTGCCCACACG 866
Qy       2535 CCTCCGCTAGTTACAGTTGTTAAAGGGCGGGTTGAGGAGCCAACAAGGTGTGAAAAGCCA 2594
Db        867 ACCCGGGTGGAGA---TGGACCGCGGCCGCATTGCAGAGCCCAGTGTGTGCGGGCG---- 919

Qy       2595 GAATGTGCAGCACGGAATGCTATGTCTCTTATTCACAATCGATGCACTTTTGCAAATAAG 2654
Db        920 --CTGCCACACCACCCACAGCATGGCACTCATCCACAACCGCTCCCTCTTCTCTGACAAG 977
Qy       2655 CAGATAGTGCGTCTTCAAGAAACTCCAGATGCCATTCCTGAAGGAGAGACTCCACACACA 2714
Db        978 CAGATGATCAAGCTTCAGGAGTCTCCGGAAGACATGCCTGCAGGGCAGACACCACACACA 1037
Qy       2715 GTCAGCATGTGTTTATACAACACTATGGTTGATGCTGTGAAGCCTGGAGATCGTATTGAG 2774
Db       1038 GTTATCCTGTTTGCTCACAATGATCTCGTTGACAAGGTCCAGCCTGGGGACAGAGTGAAT 1097
Qy       2775 GTAACAGGAGTTTTCAAGGCCATGGCAGTTCGAGTTGGTCCGAATCAACGAACATTACGA 2834
Db       1098 GTTACAGGCATCTATCGAGCTGTGCCTATTCGAGTCAATCCAAGAGTGAGTAATGTGAAG 1157
Qy       2835 GCATTGTATAAGACCTACATCGATTGCGTGCACGTCAAGAAGTCTGACAGGGGTCGACTG 2894
Db       1158 TCTGTCTACAAAACCCACATTGATGTCATTCATTATCGGAAAACGGATGCAAAACGTCTG 1217
Qy       2895 CAAACTGAAGATCCTATGGAGATGGATAAGGAGAATGATATGTATGCTGGGTATCATGAA 2954
Db       1218 CA-------------------------------------------TGGCCTTGATGAA 1232
Qy       2955 AGTGATACTTCAGAAGCTGCTAATGAAGCAAAGATTCAAAAACTTAAAGAGCTGTCCAAG 3014
Db       1233 GAAGCAGAACAGAAACTTTTTTCAGAGAAACGTGTGGAATTGCTTAAGGAACTTTCCAGG 1292
Qy       3015 CTCCCGGGCATTTATGATAGACTTTCAAGGTCGCTGGCTCCAAGCATTTGGGAGCTTGAA 3074
Db       1293 AAACCAGACATTTATGAGAGGCTTGCTTCAGCCTTGGCTCCAAGCATTTATGAACATGAA 1352
Qy       3075 GATATTAAAAAGGGTCTTCTTTGCCAGCTCTTTGGTGGGAAGGCTAAGAAAATT------ 3128
Db       1353 GATATAAAGAAGGGAATTTTGCTTCAGCTCTTTGGCGGGACAAGGAAGGATTTTAGTCAC 1412
Qy       3129 CCATCTGGAGCATCTTTCCGAGGTGACATCAATGTTTTACTTGTTGGGGACCCTGGTACC 3188
Db       1413 ACTGGAAGGGGCAAATTTCGGGCTGAGATCAACATCTTGCTGTGTGGCGACCCTGGTACC 1472
Qy       3189 AGTAAATCTCAGCTGCTTCAGTATGTGCACAAGATAGCTCCTCGTGGAATCTACACTAGT 3248
Db       1473 AGCAAGTCCCAGCTGCTGCAGTACGTGTACAACCTCGTCCCCAGGGGCCAGTACACGTCT 1532
Qy       3249 GGGCGAGGAAGTTCGGCGGTTGGGCTGACAGCGTATGTAACGAAGGATCCAGAAACTCGA 3308
Db       1533 GGGAAGGGCTCCAGTGCAGTTGGCCTCACTGCGTACGTAATGAAAGACCCTGAGACAAGG 1592
Qy       3309 GAGACGGTATTGGAGAGCGGAGCTTTGGTTCTTAGTGATCGTGGGATATGCTGTATCGAT 3368
Db       1593 CAGCTGGTCCTGCAGACAGGTGCTCTTGTCCTGAGTGACAACGGCATCTGCTGTATCGAT 1652
Qy       3369 GAGTTCGACAAAATGTCTGATAATGCCCGAAGCATGCTTCATGAGGTAATGGAGCAACAA 3428
Db       1653 GAGTTCGACAAGATGAATGAAAGTACAAGATCGGTATTGCATGAAGTCATGGAACAGCAG 1712
Qy       3429 ACGGTATCTGTAGCCAAAGGGGGTATCATTGCCTCGCTGAACGCTCGGACGTCTGTCCTT 3488
Db       1713 ACTCTGTCCATTGCAAAGGCTGGGATCATCTGTCAGCTCAATGCGCGCACCTCTGTCCTG 1772
Qy       3489 GCATGTGCAAATCCTAGTGGGTCCCGATACAATGCGCGCCTTTCTGTGATTGATAACATC 3548
Db       1773 GCAGCAGCAAATCCCATTGAGTCTCAGTGGAATCCTAAAAAAACAACCATTGAAAACATC 1832
Qy       3549 CAGCTTCCTCCAACTCTACTTTCTAGATTTGATTTAATTTACTTAATGCTCGACAAACCA 3608
Db       1833 CAGCTGCCTCATACTTTATTATCAAGGTTTGATTTGATCTTCCTCATGCTGGACCCTCAG 1892
Qy       3609 GACGAGCAAAACGATCGTCGTCTCGCCAGGCATCTCGTGGCTTTACACTATGAAAACTAT 3668
Db       1893 GACGAAGCCTATGACAGGCGTCTGGCTCACCACCTGGTCGCACTGTACTACCAGAGCGAG 1952
Qy       3669 GAAGTTTCAAAGCAGGACGCCTTAGATCTACAAACACTTACCGCGTATATCACCTATGCT 3728
```

```
Db       1953 GAGCAGGCAGAGGAGGAGCTCCTGGACATGGCGGTGCTAAAGGACTACATTGCCTACGCG 2012

Qy       3729 CGTCAGCATGTACATCCTACATTAAGTGATGAAGCTGCTGAAGATTTGATTAATGGCTAT 3788

Db       2013 CACAGCACCATCATGCCGCGGCTAAGTGAGGAAGCCAGCCAGGCTCTCATCGAGGCTTAT 2072

Qy       3789 GTTGAGATGCGCCAAAAGGGCAACTTTCCTGGAAGCAGTAAAAAGGTGATAACAGCCACA 3848

Db       2073 GTAGACATGAGGAAGATTGGCAGTAGCCGGGGAA------------TGGTTTCTGCATAC 2120

Qy       3849 CCTCGGCAACTCGAAAGTATGATTCGTATCAGTGAAGCCCTAGCTCGAATGAGATTTTCT 3908

Db       2121 CCTCGACAGCTAGAGTCATTAATCCGCTTAGCAGAAGCCCATGCTAAAGTAAGATTGTCT 2180

Qy       3909 GAAGTGGTAGAGAAAGTTGATGCAGCAGAAGCTGTGCGCCTTTTAGACGTCGCTTTGCAG 3968

Db       2181 AACAAAGTTGAAGCCATTGATGTGGAAGAGGCCAAACGCCTCCATCGGGAAGCTCTGAAG 2240

Qy       3969 CAATCTGCTACTGATCATGCAACAGGTACGATAGACATGGATCTTATCACGACTGGAGTG 4028

Db       2241 CAGTCTGCAACTGATCCCCGGACTGGCATCGTGGACATATCTATTCTTACTACGGGGATG 2300

Qy       4029 TCGGCCAGCGAGCGTATTCGTCGGGCCAACTTGCTAGCTGCTCTGCGAGAGCTTATAGCA 4088

Db       2301 AGTGCCACCTCTCGTAAACGGAAAGAAGAATTAGCTGAAGCATTGAAAAAGCTTATTTTA 2360

Qy       4089 GATAAAATTTCACCTGGCAGCTCCTCTGGCTTGAAGACCAGTCAGCTTCTTGAGGATATC 4148

Db       2361 TCTAA---------GGGCAAAACACCAGCTCTAAAATACCAGCAACTTTTTGAAGATATT 2411

Qy       4149 CGGAGCCAAAGCAGTGTGGACGTTAGTTTGCAGGATATTAAAAATGCTCTGGGTAGCCTC 4208

Db       2412 CGGGGACAATCTGACATAGCAATTACTAAAGATATGTTTGAAGAAGCACTGCGTGCCCTG 2471

Qy       4209 CAAGGAGAAGGCTTTCTTACTGTCCATGGTGACATAGT 4246

Db       2472 GCAGATGATGATTTCCTGACAGTGACTGGGAAGACCGT 2509
```

RESULT 10
BM410852

```
LOCUS       BM410852                775 bp    mRNA    linear   EST 22-JAN-2002
DEFINITION  EST585179 tomato breaker fruit Lycopersicon esculentum cDNA clone
            cLEG54H12 5' end, mRNA sequence.
ACCESSION   BM410852
VERSION     BM410852.1  GI:18262482
KEYWORDS    EST.
SOURCE      Lycopersicon esculentum (Solanum lycopersicum)
  ORGANISM  Lycopersicon esculentum
            Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
            Spermatophyta; Magnoliophyta; eudicotyledons; core eudicotyledons;
            asterids; lamiids; Solanales; Solanaceae; Solanum; Lycopersicon.
REFERENCE   1  (bases 1 to 775)
  AUTHORS   Alcala,J., Vrebalov,J., White,R., Vision,T., Karamycheva,S.A.,
            Tsai,J., Bougri,O., Kirkness,E., Utterback,T., Van Aken,S.,
            Ronning,C.M., Fraser,C.M., Martin,G.B., Tanksley,S.D. and
            Giovannoni,J.
  TITLE     Generation of ESTs from tomato fruit tissue, breaker stage (2002)
  JOURNAL   Unpublished (2002)
COMMENT     Contact: CUGI
            Clemson University Genomics Institute
            Clemson University
            100 Jordan Hall, Clemson, SC 29634, USA
            Email: http://www.genome.clemson.edu/orders/index.html
            This clone is available through the Clemson University Genomics
            Institute
            Seq primer: T3.
FEATURES             Location/Qualifiers
     source          1..775
                     /organism="Lycopersicon esculentum"
                     /mol_type="mRNA"
                     /cultivar="TA496"
                     /db_xref="taxon:4081"
                     /clone="cLEG54H12"
                     /tissue_type="Pericarp"
                     /dev_stage="breaker"
                     /lab_host="SOLR"
                     /clone_lib="tomato breaker fruit"
                     /note="Vector: pBluescriptSKmCUadapt; Site_1: EcoR1;
                     Site_2: XhoI; supplier: Boyce Thompson Institute;
                     sequencing: The Institute for Genomic Research.  Fruit
                     were harvested at the breaker stage (first sign of
                     lycopene accumulation on the blossom end of fruit).  Fruit
                     were cut in half and the seeds and locules were discarded
                     prior to freezing the pericarp."
ORIGIN

  Query Match             9.8%;  Score 428;  DB 3;  Length 775;
  Best Local Similarity   73.6%;  Pred. No. 3.9e-120;
  Matches  545;  Conservative    0;  Mismatches  195;  Indels    0;  Gaps    0;

Qy       3321 GAGAGCGGAGCTTTGGTTCTTAGTGATCGTGGGATATGCTGTATCGATGAGTTCGACAAA 3380

Db          2 GAGAGTGGCGCTCTGGTCTTGAGTGACAGAGGGATCTGCTGTATTGACGAGTTTGACAAA 61

Qy       3381 ATGTCTGATAATGCCCGAAGCATGCTTCATGAGGTAATGGAGCAACAAACGGTATCTGTA 3440

Db         62 ATGTCTGACAGTGCCAGGAGCATGTTACATGAGGTGATGGAGCAACAAACTGTTTCAATT 121

Qy       3441 GCCAAAGGGGGTATCATTGCCTCGCTGAACGCTCGGACGTCTGTCCTTGCATGTGCAAAT 3500

Db        122 GCAAAGGCTGGAATTATCGCTTCGCTTAATGCTAGGACTTCAGTATTGGCATGTGCAAAT 181

Qy       3501 CCTAGTGGGTCCCGATACAATGCGCGCCTTTCTGTGATTGATAACATCCAGCTTCCTCCA 3560

Db        182 CCAATTGGCTCCCGTTACAATCCCCGATTATCTGTCATTGATAATATACACCTTCCACCT 241

Qy       3561 ACTCTACTTTCTAGATTTGATTTAATTTACTTAATGCTCGACAAACCAGACGAGCAAAAC 3620

Db        242 ACCCTGCTGTCTAGATTTGATTTGATATATTTAATTCTAGACAAAGCAGATGAGCAGACC 301

Qy       3621 GATCGTCGTCTCGCCAGGCATCTCGTGGCTTTACACTATGAAAACTATGAAGTTTCAAAG 3680

Db        302 GACAGGCGCCTTGCAAAGCACATAGTTGCTTTACACTCTGAGAATCCTGAGAACTCTGAG 361

Qy       3681 CAGGACGCCTTAGATCTACAAACACTTACCGCGTATATCACCTATGCTCGTCAGCATGTA 3740

Db        362 CAAGAAGTGATTGACCTTCCAACATTAGCTTCATACTTGAGCTATGCTCGAAAACATATA 421

Qy       3741 CATCCTACATTAAGTGATGAAGCTGCTGAAGATTTGATTAATGGCTATGTTGAGATGCGC 3800

Db        422 CATCCACAATTATCTGATGAAGCAGCTGAAGAATTGACTAGAGGTTATGTGGAGATGAGA 481

Qy       3801 CAAAAGGGCAACTTTCCTGGAAGCAGTAAAAAGGTGATAACAGCCACACCTCGGCAACTC 3860

Db        482 AGAAAAGGGAATTTCCCCGGTAGCAGTAAAAAGGTGATTACAGCTACACCAAGGCAGCTA 541

Qy       3861 GAAAGTATGATTCGTATCAGTGAAGCCCTAGCTCGAATGAGATTTTCTGAAGTGGTAGAG 3920

Db        542 GAGAGTTTGATACGTCTAGGTGAAGGTCTTGCTCGGATGCGTTTTTCAGAAAAGGTGGAG 601

Qy       3921 AAAGTTGATGCAGCAGAAGCTGTGCGCCTTTTAGACGTCGCTTTGCAGCAATCTGCTACT 3980

Db        602 AAAAGAGATGTAGTGGAGGCTTTTCGACTTCTAGAAGTTGCATTGCAGCAGTCTGCAACT 661

Qy       3981 GATCATGCAACAGGTACGATAGACATGGATCTTATCACGACTGGAGTGTCGGCCAGCGAG 4040

Db        662 GACCATGCTACAGGAACCATTGACATGGATCTCATCACAACTGGAGTATCTGCAAGTGAA 721

Qy       4041 CGTATTCGTCGGGCCAACTT 4060

Db        722 AGGATGAGAAGGGGAGAATT 741
```

RESULT 11

```
AY408034
LOCUS       AY408034                2357 bp    DNA     linear   GSS 12-DEC-2003
DEFINITION  Pan troglodytes HCM3089 gene, VIRTUAL TRANSCRIPT, partial sequence,
            genomic survey sequence.
ACCESSION   AY408034
VERSION     AY408034.1  GI:39764005
KEYWORDS    GSS.
SOURCE      Pan troglodytes (chimpanzee)
  ORGANISM  Pan troglodytes
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Euarchontoglires; Primates; Catarrhini;
            Hominidae; Pan.
REFERENCE   1  (bases 1 to 2357)
  AUTHORS   Clark,A.G., Glanowski,S., Nielson,R., Thomas,P., Kejariwal,A.,
            Todd,M.A., Tanenbaum,D.M., Civello,D.R., Lu,F., Murphy,B.,
            Ferriera,S., Wang,G., Zheng,X.H., White,T.J., Sninsky,J.J.,
            Adams,M.D. and Cargill,M.
  TITLE     Inferring nonneutral evolution from human-chimp-mouse orthologous
            gene trios
  JOURNAL   Science 302 (5652), 1960-1963 (2003)
   PUBMED   14671302
REFERENCE   2  (bases 1 to 2357)
  AUTHORS   Clark,A.G., Glanowski,S., Nielson,R., Thomas,P., Kejariwal,A.,
            Todd,M.A., Tanenbaum,D.M., Civello,D.R., Lu,F., Murphy,B.,
            Ferriera,S., Wang,G., Zheng,X.H., White,T.J., Sninsky,J.J.,
            Adams,M.D. and Cargill,M.
  TITLE     Direct Submission
  JOURNAL   Submitted (16-NOV-2003) Celera Genomics, 45 West Gude Drive,
            Rockville, MD 20850, USA
COMMENT     This sequence was made by sequencing genomic exons and ordering
            them based on alignment.
FEATURES             Location/Qualifiers
     source          1..2357
                     /organism="Pan troglodytes"
                     /mol_type="genomic DNA"
                     /db_xref="taxon:9598"
     gene            <1..>2357
                     /locus_tag="HCM3089"
ORIGIN

  Query Match             9.4%;  Score 408.4;  DB 10;  Length 2357;
  Best Local Similarity   50.4%;  Pred. No. 7.1e-114;
  Matches 1015;  Conservative    0;  Mismatches  922;  Indels   78;  Gaps    6;

Qy       2181 TACATCCAGATCATAGAGGAGACTGTGGAGCGTGAGGAGGATACTCTAAATATCGACATG 2240
Db        342 TACATGCAACGACTTGGGGAGATTAATGTTATTGGTGAGCCATTTTTAAATGTGAACTGT 401

Qy       2241 TCAGACATTTATGACCATGATCCTGATCTATACGCAAAAATTGTTCGATACCCACTCGAC 2300
Db        402 GAACACATCAAATCATTTGACAAAAATTTGTACAGACAACTCATCTCTTACCCACAGGAA 461

Qy       2301 ATCATCCCCCTGTTGGACACTGAGTGTCAGGAA------GTTGCTACCTCTTTACTACCA 2354
Db        462 GTTATTCCAACTTTTGACATGGCTGTCAATGAAATCTTCTTTGACCGTTACCCTGACTCA 521

Qy       2355 ACGTTTGAGAAGCATATTGAGGCCAGACCTTTCAATCTCAAAGCATCGGTGCACATGCGT 2414
Db        522 ATCTTAGAACATCAGATTCAAGTAAGACCATTCAACGCATTGAAGACTAAGAATATGAGA 581

Qy       2415 GAACTCAACCCTTCAGATATAGACAAATTGGTTTCTGTTAAAGGAATGGTTATCCGGTGC 2474
Db        582 AACCTGAATCCAGAAGACATTGACCAGCTCATCACCATCAGTGGCATGGTGATCAGGACA 641

Qy       2475 AGTTCTATCATACCTGAAATTAAGGGGGCCTTCTTCAAATGTTTAGTGTGTGGTCACTCG 2534
Db        642 TCCCAGCTGATTCCCGAGATGCAGGAGGCCTTCTTCCAGTGCCAAGTGTGTGCCCACACG 701

Qy       2535 CCTCCGCTAGTTACAGTTGTTAAAGGGCGGGTTGAGGAGCCAACAAGGTGTGAAAAGCCA 2594
Db        702 ACCCGGGTGGAGA---TGGACCGCGGCCGCATTGCAGAGCCCAGTGTGTGCG------GG 752

Qy       2595 GAATGTGCAGCACGGAATGCTATGTCTCTTATTCACAATCGATGCACTTTTGCAAATAAG 2654
Db        753 CGCTGCCACACCACCCACAGCATGGCACTCATCCACAACCGCTCCCTCTTCTCTGACAAG 812

Qy       2655 CAGATAGTGCGTCTTCAAGAAACTCCAGATGCCATTCCTGAAGGAGAGACTCCACACACA 2714
Db        813 CAGATGATCAAGCTTCAGGAGTCTCCGGAAGACATGCCTGCAGGGCAGACACCACACACA 872

Qy       2715 GTCAGCATGTGTTTATACAACACTATGGTTGATGCTGTGAAGCCTGGAGATCGTATTGAG 2774
Db        873 GTTATCCTGTTTGCTCACAATGATCTCGTTGACAAGGTCCAGCCTGGGGACAGAGTGAAT 932

Qy       2775 GTAACAGGAGTTTTCAAGGCCATGGCAGTTCGAGTTGGTCCGAATCAACGAACATTACGA 2834
Db        933 GTTACAGGCATCTATCGAGCTGTGCCTATTCGAGTCAATCCAAGAGTGAGTAATGTGAAG 992

Qy       2835 GCATTGTATAAGACCTACATCGATTGCGTGCACGTCAAGAAGTCTGACAGGGGTCGACTG 2894
Db        993 TCTGTCTACAAAACCCACATTGATGTCATTCATTATCGGAAAACGGATGCAAAACGTCTG 1052

Qy       2895 CAAACTGAAGATCCTATGGAGATGGATAAGGAGAATGATATGTATGCTGGGTATCATGAA 2954
Db       1053 CA-------------------------------------------TGGCCTTGATGAA 1067

Qy       2955 AGTGATACTTCAGAAGCTGCTAATGAAGCAAAGATTCAAAAACTTAAAGAGCTGTCCAAG 3014
Db       1068 GAAGCAGAACAGAAACTTTTTTCAGAGAAACGTGTGGAATTGCTTAAGGAACTTTCCAGG 1127

Qy       3015 CTCCCGGGCATTTATGATAGACTTTCAAGGTCGCTGGCTCCAAGCATTTGGGAGCTTGAA 3074
Db       1128 AAACCAGACATTTATGAGAGGCTTGCTTCAGCCTTGGCTCCAAGCATTTATGAACATGAA 1187

Qy       3075 GATATTAAAAAGGGTCTTCTTTGCCAGCTCTTTGGTGGGAAGGCTAAGAAAATT------ 3128
Db       1188 GATATAAAGAAGGGAATTTTGCTTCAGCTCTTTGGCGGGACAAGGAAGGATTTTAGTCAC 1247

Qy       3129 CCATCTGGAGCATCTTTCCGAGGTGACATCAATGTTTTACTTGTTGGGGACCCTGGTACC 3188
Db       1248 ACTGGAAGGGGCAAATTTCGGGCTGAGATCAACATCTTGCTGTGTGGCGACCCTGGTACC 1307

Qy       3189 AGTAAATCTCAGCTGCTTCAGTATGTGCACAAGATAGCTCCTCGTGGAATCTACACTAGT 3248
Db       1308 AGCAAGTCCCAGCTGCTGCAGTATGTGTACAACCTCGTCCCCAGGGGCCAGTACACGTCT 1367

Qy       3249 GGGCGAGGAAGTTCGGCGGTTGGGCTGACAGCGTATGTAACGAAGGATCCAGAAACTCGA 3308
Db       1368 GGGAAGGGCTCCAGTGCAGTTGGCCTCACTGCGTACGTAATGAAAGACCCTGAGACAAGG 1427

Qy       3309 GAGACGGTATTGGAGAGCGGAGCTTTGGTTCTTAGTGATCGTGGGATATGCTGTATCGAT 3368
Db       1428 CAGCTGGTCCTGCAGACAGGTGCTCTTGTCCTGAGTGACAACGGCATCTGCTGTATCGAT 1487

Qy       3369 GAGTTCGACAAAATGTCTGATAATGCCCGAAGCATGCTTCATGAGGTAATGGAGCAACAA 3428
Db       1488 GAGTTCGACAAGATGAATGAAAGTACAAGATCGGTATTGCATGAAGTCATGGAACAGCAG 1547

Qy       3429 ACGGTATCTGTAGCCAAAGGGGGTATCATTGCCTCGCTGAACGCTCGGACGTCTGTCCTT 3488
Db       1548 ACTCTGTCCATTGCAAAGGCTGGGATCATCTGTCAGCTCAATGCGCGCACCTCTGTCCTG 1607

Qy       3489 GCATGTGCAAATCCTAGTGGGTCCCGATACAATGCGCGCCTTTCTGTGATTGATAACATC 3548
Db       1608 GCAGCAGCAAATCCCATTGAGTCTCAGTGGAATCCTAAAAAAACAACCATTGAAAACATC 1667

Qy       3549 CAGCTTCCTCCAACTCTACTTTCTAGATTTGATTTAATTTACTTAATGCTCGACAAACCA 3608
Db       1668 CAGCTGCCTCATACTTTATTATCAAGGTTTGATTTGATCTTCCTCATGCTGGACCCTCAG 1727

Qy       3609 GACGAGCAAAACGATCGTCGTCTCGCCAGGCATCTCGTGGCTTTACACTATGAAAACTAT 3668
Db       1728 GACGAAGCCTATGACAGGCGTCTGGCTCACCACCTGGTCGCACTGTACTACCAGAGCGAG 1787

Qy       3669 GAAGTTTCAAAGCAGGACGCCTTAGATCTACAAACACTTACCGCGTATATCACCTATGCT 3728
```

```
Db        1788 GAGCAGGCAGAGGAGGAGCTCCTGGACATGGCGGTGCTAAAGGACTACATTGCCTACGCA 1847

Qy        3729 CGTCAGCATGTACATCCTACATTAAGTGATGAAGCTGCTGAAGATTTGATTAATGGCTAT 3788

Db        1848 CACAGCACCATCATGCCGCGGCTAAGTGAGGAAGCCAGCCAGGCTCTCATCGAGGCTTAT 1907

Qy        3789 GTTGAGATGCGCCAAAAGGGCAACTTTCCTGGAAGCAGTAAAAGGTGATAACAGCCACA 3848

Db        1908 GTAGACATGAGGAAGATTGGCAGTAGCCGGGGAA-----------TGGTTTCTGCATAC 1955

Qy        3849 CCTCGGCAACTCGAAAGTATGATTCGTATCAGTGAAGCCCTAGCTCGAATGAGATTTTCT 3908

Db        1956 CCTCGACAGCTAGAGTCATTAATCCGCTTAGCAGAAGCCCATGCTAAAGTAAGATTGTCT 2015

Qy        3909 GAAGTGGTAGAGAAAGTTGATGCAGCAGAAGCTGTGCGCCTTTTAGACGTCGCTTTGCAG 3968

Db        2016 AACAAAGTTGAAGCCATTGATGTGGAAGAGGCCAAACGCCTCCANNNNNNNGCTCTGAAG 2075

Qy        3969 CAATCTGCTACTGATCATGCAACAGGTACGATAGACATGGATCTTATCACGACTGGAGTG 4028

Db        2076 CAGTCTGCAACTGATCCCCGGACTGGCATCGTGGACATATCTATTCTTACTACGGNNNNN 2135

Qy        4029 TCGGCCAGCGAGCGTATTCGTCGGGCCAACTTGCTAGCTGCTCTGCGAGAGCTTATAGCA 4088

Db        2136 NNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNN 2195

Qy        4089 GATAAAATTTCACCTGGCAGCTCCTCTGGCTTGAAGACCAGTCAGCTTCTTGAGGATATC 4148

Db        2196 NNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNN 2255

Qy        4149 CGGAGCCAAAGCAGTGTGGACGTTAGTTTGCAGGA 4183

Db        2256 NNNNNNNNNGCAATTACTAAAGATATGTTTGAAGA 2290

RESULT 12
CF449139
LOCUS       CF449139               798 bp    mRNA    linear   EST 04-SEP-2003
DEFINITION  EST685484 normalized cDNA library of onion Allium cepa cDNA clone
            ACABW59, mRNA sequence.
ACCESSION   CF449139
VERSION     CF449139.1  GI:34471841
KEYWORDS    EST.
SOURCE      Allium cepa (onion)
  ORGANISM  Allium cepa
            Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
            Spermatophyta; Magnoliophyta; Liliopsida; Asparagales; Alliaceae;
            Allium.
REFERENCE   1  (bases 1 to 798)
  AUTHORS   Havey,M.J., Cheung,F., Van Aken,S., Utterback,T. and Town,C.D.
  TITLE     Expressed Sequence Tags from a normalized library of mixed onion
            tissues (Allium cepa)
  JOURNAL   Unpublished (2003)
COMMENT     Contact: Havey MJ
            Department of Horticulture
            USDA-ARS and University of Wisconsin
            1575 Linden Drive, Madison, WI 53706, USA
            Tel: 608-262-1830
            Fax: 608-262-4743
            Email: mjhavey@facstaff.wisc.edu
            TIGR sequence name ACABW59TR. For more information:
            http://haveylab.hort.wisc.edu
            Seq primer: CAG GAA ACA GCT ATG ACC.
FEATURES             Location/Qualifiers
     source          1..798
                     /organism="Allium cepa"
                     /mol_type="mRNA"
                     /cultivar="Red Creole(bulbs), unknown(callus), Ebano &
                     Texas Legend(roots)"
                     /db_xref="taxon:4679"
                     /clone="ACABW59"
                     /tissue_type="Callus, roots, and young bulbs"
                     /clone_lib="normalized cDNA library of onion"
                     /note="Vector: pCMVSport6.1-ccdb (Invitrogen); Site_1:
                     EcoRV (5'); Site_2: NotI (3'); Equal molar amounts of mRNA
                     from callus, roots, and young bulbs were combined to
                     synthesize the library. Normalization to enrich for
                     low-copy transcripts was performed by proprietary
                     techniques of Invitrogen."
ORIGIN

  Query Match             9.3%;  Score 404.2;  DB 6;  Length 798;
  Best Local Similarity   72.5%;  Pred. No. 9.3e-113;
  Matches  523;  Conservative    0;  Mismatches  198;  Indels    0;  Gaps    0;

Qy        3376 ACAAAATGTCTGATAATGCCCGAAGCATGCTTCATGAGGTAATGGAGCAACAAACGGTAT 3435

Db           1 ATAAGATGTCTGATAATGCTCGTAGTATGTTACATGAGGTTATGGAACAACAAACTGTAT 60

Qy        3436 CTGTAGCCAAAGGGGGTATCATTGCCTCGCTGAACGCTCGGACGTCTGTCCTTGCATGTG 3495

Db          61 CAATTGCAAAAGCTGGTATAATCGCTTCTCTGAATGCACGTACTTCTGTTTTAGCCTGTG 120

Qy        3496 CAAATCCTAGTGGGTCCCGATACAATGCGCGCCTTTCTGTGATTGATAACATCCAGCTTC 3555

Db         121 CAAATCCAATTGGGTCTCGTTATAACCCTCGAATGTCTGTGATCGAGAACATTCACCTTC 180

Qy        3556 CTCCAACTCTACTTTCTAGATTTGATTTAATTTACTTAATGCTCGACAAACCAGACGAGC 3615

Db         181 CTCCAACTTTACTATCTAGATTTGATTTGATCTACTTGATACTGGACAAAGCAGACGAAC 240

Qy        3616 AAAACGATCGTCGTCTCGCCAGGCATCTCGTGGCTTTACACTATGAAAACTATGAAGTTT 3675

Db         241 AAATGGATAGACGGTTGGCTAAGCATATTGTTTCCTTGCATTTTGAGAACCCTGAGAGTG 300

Qy        3676 CAAAGCAGGACGCCTTAGATCTACAAACACTTACCGCGTATATCACCTATGCTCGTCAGC 3735

Db         301 TGGTACAAGATGTCTTGGATCTTCCTACGTTAACTTCATACGTTAGCTATGCAAGAAAGC 360

Qy        3736 ATGTACATCCTACATTAAGTGATGAAGCTGCTGAAGATTTGATTAATGGCTATGTTGAGA 3795

Db         361 ATATTCATCCCAAGTTATCCGATGAAGCTGCTGAAGAACTGACTCGTGGCTATGTTGAAA 420

Qy        3796 TGCGCCAAAAGGGCAACTTTCCTGGAAGCAGTAAAAAGGTGATAACAGCCACACCTCGGC 3855

Db         421 TGAGGAAAAGAGGAAACACTCCTGGTAGCAGTAAAAAGGTGATTACAGCCACTGCTAGGC 480

Qy        3856 AACTCGAAAGTATGATTCGTATCAGTGAAGCCCTAGCTCGAATGAGATTTTCTGAAGTGG 3915

Db         481 AAATTGAGAGTTTGATTCGGCTGAGTGAAGCACTGGCTCGAATGCGATTCTCAGAATTGG 540

Qy        3916 TAGAGAAAGTTGATGCAGCAGAAGCTGTGCGCCTTTTAGACGTCGCTTTGCAGCAATCTG 3975

Db         541 TAGAAGCAGGTGATGTGCAAGAAGCGTTTAGGCTACTTGATGTTGCAATGCAGCAGTCTG 600

Qy        3976 CTACTGATCATGCAACAGGTACGATAGACATGGATCTTATCACGACTGGAGTGTCGGCCA 4035

Db         601 CAACTGATCATGCCACTGGTACGATTGACATGGATCTTATAATGACTGGAGTTTCCGCTA 660

Qy        4036 GCGAGCGTATTCGTCGGGCCAACTTGCTAGCTGCTCTGCGAGAGCTTATAGCAGATAAAA 4095

Db         661 GTGAAAGAATGAGACGTGAAAATCTAGTTTCAGCAACTCGTGACCTTATAATGGATAAAA 720

Qy        4096 T 4096

Db         721 T 721

RESULT 13
DR917805
LOCUS       DR917805               709 bp    mRNA    linear   EST 02-AUG-2005
DEFINITION  EST1109344 Aquilegia cDNA library Aquilegia formosa x Aquilegia
            pubescens cDNA clone C01LX60, mRNA sequence.
ACCESSION   DR917805
VERSION     DR917805.1  GI:71687168
KEYWORDS    EST.
```

```
SOURCE      Aquilegia formosa x Aquilegia pubescens
  ORGANISM  Aquilegia formosa x Aquilegia pubescens
            Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
            Spermatophyta; Magnoliophyta; eudicotyledons; Ranunculales;
            Ranunculaceae; Aquilegia.
REFERENCE   1  (bases 1 to 709)
  AUTHORS   Hodges,S.A., Rensink,W., Buell,C.R., Borevitz,J., Kramer,E.,
            Nordborg,M. and Tomkins,J.
  TITLE     Generation of ESTs from Aquilegia
  JOURNAL   Unpublished (2005)
COMMENT     Other_ESTs: EST1109343
            Contact: Scott Hodges
            Department of Ecology, Evolution and Marine Biology
            University of California, Santa Barbara
            Santa Barbara, CA 93106, USA
            Tel: 805 893 7813
            Fax: 805 893 4724
            Email: hodges@lifesci.ucsb.edu
            Seq primer: M13 Reverse.
FEATURES             Location/Qualifiers
     source          1..709
                     /organism="Aquilegia formosa x Aquilegia pubescens"
                     /mol_type="mRNA"
                     /db_xref="taxon:338618"
                     /clone="C01LX60"
                     /tissue_type="mixed shoot and floral apical meristems,
                     flower buds, leaves and roots"
                     /lab_host="DH10B T1 (T1 and T5 phage resistance)"
                     /clone_lib="Aquilegia cDNA library"
                     /note="Vector: pCMV SPORT6.1; Site_1: EcoRI; Site_2: NotI;
                     F2, F3, and F4 lines of Aquilegia formosa X A. pubescens
                     were grown from seed in greenhouses at UC Santa Barbara.
                     From these plants three sets of tissue were collected: 1)
                     Small flower buds (<10 mm) and very young inflorescences
                     (71 & 29% by weight respectively), 2) Medium (7-20 mm) and
                     large (at or near anthesis) flower buds (65 & 35% by
                     weight respectively) and 3) Shoot apical meristems. A
                     fourth set of tissue was collected from plants of A.
                     formosa. These plants were grown from seed in sand and at
                     approximately 1 month root tissue and leaf tissue of
                     various developmental stages were collected (84 & 16% by
                     weight respectively). Total RNA was extracted from each
                     set of tissue and pooled in the following proportions:
                     1.5X from sets 1 & 2, 1X from sets 3 & 4. From the pooled
                     total RNA, mRNA was extracted and enriched for full-length
                     messages and then normalized with proprietary methods by
                     Invitrogen."
ORIGIN

  Query Match             9.1%;  Score 394.4;  DB 8;  Length 709;
  Best Local Similarity   72.3%;  Pred. No. 9.6e-110;
  Matches  512;  Conservative    0;  Mismatches  196;  Indels    0;  Gaps    0;

Qy       3105 TTTGGTGGGAAGGCTAAGAAAATTCCATCTGGAGCATCTTTCCGAGGTGACATCAATGTT 3164
Db          2 TTTGGTGGGAATGCTTTGAAGCTGCCTACAGGTGCTACCTTTCGTGGTGACATCAACATT 61

Qy       3165 TTACTTGTTGGGGACCCTGGTACCAGTAAATCTCAGCTGCTTCAGTATGTGCACAAGATA 3224
Db         62 CTGCTCGTCGGGGATCCTGGGACTAGCAAGTCCCAGCTGCTCCAATATATACACAAGCTT 121

Qy       3225 GCTCCTCGTGGAATCTACACTAGTGGGCGAGGAAGTTCGGCGGTTGGGCTGACAGCGTAT 3284
Db        122 TCCCCTCGGGGAATATATACCAGTGGAAGAGGGAGTTCAGCTGTGGGGTTAACTGCGTAT 181

Qy       3285 GTAACGAAGGATCCAGAAACTCGAGAGACGGTATTGGAGAGCGGAGCTTTGGTTCTTAGT 3344
Db        182 GTTACTAAGGACCCAGAAACGGGTGAAACAGTATTAGAGAGTGGAGCCCTGGTTCTGAGC 241

Qy       3345 GATCGTGGGATATGCTGTATCGATGAGTTCGACAAAATGTCTGATAATGCCCGAAGCATG 3404
Db        242 GATAGAGGCATCTGCTGCATTGATGAATTTGACAAAATGTCTGAGAATGCAAGGAGTATG 301

Qy       3405 CTTCATGAGGTAATGGAGCAACAAACGGTATCTGTAGCCAAAGGGGGTATCATTGCCTCG 3464
Db        302 TTGCATGAGGTTATGGAACAGCAAACTGTTTCAATAGCCAAAGCAGGGATAATTGCTTCT 361

Qy       3465 CTGAACGCTCGGACGTCTGTCCTTGCATGTGCAAATCCTAGTGGGTCCCGATACAATGCG 3524
Db        362 CTTAATGCCAGGACATCAGTACTAGCTTGTGCAAACCCTAGTGGGTCCCGCTATAATCCT 421

Qy       3525 CGCCTTTCTGTGATTGATAACATCCAGCTTCCTCCAACTCTACTTTCTAGATTTGATTTA 3584
Db        422 CGTTTATCTGTAATCGACAATATCCACCTTCCTCCAACATTACTTTCCAGGTTTGATCTT 481

Qy       3585 ATTTACTTAATGCTCGACAAACCAGACGAGCAAAACGATCGTCGTCTCGCCAGGCATCTC 3644
Db        482 ATATACTTAATGCTTGACAAGGCTGATGAACATATGGATAGGCGTCTTGCTAAACATATT 541

Qy       3645 GTGGCTTTACACTATGAAAACTATGAAGTTTCAAAGCAGGACGCCTTAGATCTACAAACA 3704
Db        542 GTGGCACTTCATTTTGAGAATCCTGAGACTGTGCAGCAAGATGTCTTAGACATACCAACA 601

Qy       3705 CTTACCGCGTATATCACCTATGCTCGTCAGCATGTACATCCTACATTAAGTGATGAAGCT 3764
Db        602 TTGACTGCATATTTGAGTTATGCCAGAAGGCACATTCGCCCAAAATTGTCAGATGAAGCT 661

Qy       3765 GCTGAAGATTTGATTAATGGCTATGTTGAGATGCGCCAAAAGGGCAAC 3812
Db        662 GCAGAAGAATTGACTAGAAGATATGTTGAGATGAGGAGAAGAGGAAAC 709


RESULT 14
AK089999
LOCUS       AK089999                3285 bp    mRNA    linear   HTC 03-APR-2004
DEFINITION  Mus musculus brain CRL-1443 BC3H1 cDNA, RIKEN full-length enriched
            library, clone:G430050N09 product:mini chromosome maintenance
            deficient 4 homolog (S. cerevisiae), full insert sequence.
ACCESSION   AK089999
VERSION     AK089999.1  GI:26354818
KEYWORDS    HTC; CAP trapper.
SOURCE      Mus musculus (house mouse)
  ORGANISM  Mus musculus
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Euarchontoglires; Glires; Rodentia;
            Sciurognathi; Muridae; Murinae; Mus.
REFERENCE   1
  AUTHORS   Carninci,P. and Hayashizaki,Y.
  TITLE     High-efficiency full-length cDNA cloning
  JOURNAL   Meth. Enzymol. 303, 19-44 (1999)
   PUBMED   10349636
REFERENCE   2
  AUTHORS   Carninci,P., Shibata,Y., Hayatsu,N., Sugahara,Y., Shibata,K.,
            Itoh,M., Konno,H., Okazaki,Y., Muramatsu,M. and Hayashizaki,Y.
  TITLE     Normalization and subtraction of cap-trapper-selected cDNAs to
            prepare full-length cDNA libraries for rapid discovery of new genes
  JOURNAL   Genome Res. 10 (10), 1617-1630 (2000)
   PUBMED   11042159
REFERENCE   3
  AUTHORS   Shibata,K., Itoh,M., Aizawa,K., Nagaoka,S., Sasaki,N., Carninci,P.,
            Konno,H., Akiyama,J., Nishi,K., Kitsunai,T., Tashiro,H., Itoh,M.,
            Sumi,N., Ishii,Y., Nakamura,S., Hazama,M., Nishine,T., Harada,A.,
            Yamamoto,R., Matsumoto,H., Sakaguchi,S., Ikegami,T., Kashiwagi,K.,
            Fujiwake,S., Inoue,K., Togawa,Y., Izawa,M., Ohara,E., Watahiki,M.,
            Yoneda,Y., Ishikawa,T., Ozawa,K., Tanaka,T., Matsuura,S., Kawai,J.,
            Okazaki,Y., Muramatsu,M., Inoue,Y., Kira,A. and Hayashizaki,Y.
  TITLE     RIKEN integrated sequence analysis (RISA) system--384-format
            sequencing pipeline with 384 multicapillary sequencer
  JOURNAL   Genome Res. 10 (11), 1757-1771 (2000)
   PUBMED   11076861
REFERENCE   4
  AUTHORS   The RIKEN Genome Exploration Research Group Phase II Team and the
            FANTOM Consortium.
  TITLE     Functional annotation of a full-length mouse cDNA collection
  JOURNAL   Nature 409, 685-690 (2001)
REFERENCE   5
```

```
AUTHORS     The FANTOM Consortium and the RIKEN Genome Exploration Research
            Group Phase I & II Team.
TITLE       Analysis of the mouse transcriptome based on functional annotation
            of 60,770 full-length cDNAs
JOURNAL     Nature 420, 563-573 (2002)
REFERENCE   6  (bases 1 to 3285)
AUTHORS     Adachi,J., Aizawa,K., Akimura,T., Arakawa,T., Bono,H., Carninci,P.,
            Fukuda,S., Furuno,M., Hanagaki,T., Hara,A., Hashizume,W.,
            Hayashida,K., Hayatsu,N., Hiramoto,K., Hiraoka,T., Hirozane,T.,
            Hori,F., Imotani,K., Ishii,Y., Itoh,M., Kagawa,I., Kasukawa,T.,
            Katoh,H., Kawai,J., Kojima,Y., Kondo,S., Konno,H., Kouda,M.,
            Koya,S., Kurihara,C., Matsuyama,T., Miyazaki,A., Murata,M.,
            Nakamura,M., Nishi,K., Nomura,K., Numazaki,R., Ohno,M., Ohsato,N.,
            Okazaki,Y., Saito,R., Saitoh,H., Sakai,C., Sakai,K., Sakazume,N.,
            Sano,H., Sasaki,D., Shibata,K., Shinagawa,A., Shiraki,T.,
            Sogabe,Y., Tagami,M., Tagawa,A., Takahashi,F., Takaku-Akahira,S.,
            Takeda,Y., Tanaka,T., Tomaru,A., Toya,T., Yasunishi,A.,
            Muramatsu,M. and Hayashizaki,Y.
TITLE       Direct Submission
JOURNAL     Submitted (16-APR-2002) Yoshihide Hayashizaki, The Institute of
            Physical and Chemical Research (RIKEN), Laboratory for Genome
            Exploration Research Group, RIKEN Genomic Sciences Center (GSC),
            RIKEN Yokohama Institute; 1-7-22 Suehiro-cho, Tsurumi-ku, Yokohama,
            Kanagawa 230-0045, Japan (E-mail:genome-res@gsc.riken.jp,
            URL:http://genome.gsc.riken.jp/, Tel:81-45-503-9222,
            Fax:81-45-503-9216)
COMMENT     cDNA library was prepared and sequenced in Mouse Genome
            Encyclopedia Project of Genome Exploration Research Group in Riken
            Genomic Sciences Center and Genome Science Laboratory in RIKEN.
            Division of Experimental Animal Research in Riken contributed to
            prepare mouse tissues.
            Please visit our web site for further details.
            URL:http://genome.gsc.riken.jp/
            URL:http://fantom.gsc.riken.jp/.
FEATURES             Location/Qualifiers
     source          1..3285
                     /organism="Mus musculus"
                     /mol_type="mRNA"
                     /strain="C3H"
                     /db_xref="FANTOM_DB:G430050N09"
                     /db_xref="taxon:10090"
                     /clone="G430050N09"
                     /cell_line="CRL-1443 BC3H1"
                     /tissue_type="brain"
                     /clone_lib="RIKEN full-length enriched mouse cDNA library"
     CDS             125..2713
                     /note="unnamed protein product; mini chromosome
                     maintenance deficient 4 homolog (S. cerevisiae)
                     (MGD|MGI:103199, GB|NM_008565, evidence: BLASTN, 100%,
                     match=3221)
                     putative"
                     /codon_start=1
                     /protein_id="BAC41036.1"
                     /db_xref="GI:26354819"
                     /translation="MSSPASTPSRRSSRRGRVTPTQSLRSEESRSSPNRRRRGEDSST
                     GELLPMPTSPGADLQSPPAQNALFSSPPQMHSLAIPLDFDVSSPLTYGTPSSRVEGTP
                     RSGVRGTPVRQRPDLGSARKGLQVDLQSDGAAAEDIVPSEQSLGQKLVIWGTDVNVAT
                     CKENFQRFLQCFTDPLAQEEENVGIDITQPLYMQQLGEINITGEPFLNVNCEHIKSFS
                     KNLYRQLISYPQEVIPTFDMAVNEIFFDRYPDSILEHQIQVRPFNALKTKSMRNLNPE
                     DIDQLITISGMVIRTSQLIPEMQEAFFQCQVCAHTTRVEIDRGRIAEPCSCVHCHTTH
                     SMALIHNRSFFSDKQMIKLQESPEDMPAGQTPHTIVLFAHNDLVDKVQPGDRVNVTGI
                     YRAVPIRVNPRVSNVKSVYKTHIDVIHYRKTDAKRLHGLDEEAEQKLFSEKRVKLLKE
                     LSRKPDIYERLASALAPSIYEHEDIKKGILLQLFGGTRKDFSHTGRGKFRAEINILLC
                     GDPGTSKSQLLQYVYNLVPRGQYTSGKGSSAVGLTAYVMKDPETRQLVLQTGALVLSD
                     NGICCIDKFDKMNESTRSVLHEVMEQQTLSIAKAGIICQLNARTSVLAAANPIESQWN
                     PKKTTIENIQLPHTLLSRFDLIFLMLDPQDEAYDRRLAHHLVSLYYQSEEQVEEEFLD
                     MAVLKDYIAYAHSTIMPRLSEEASQALIEAYVNMRKIGSSRGMVSAYPRQLESLIRLA
                     EAHAKVRFSNKVEAIDVEEAKRLHREALKQSATDPRTGMVDISILTTGMSATSRKRKE
                     ELAEALRKLILSKGKTPALKYQQLFEDIRGQSDTAITKDMFEEALRALADDDFLTVTG
                     KTVRLL"
     polyA_signal    3268..3273
                     /note="putative"
     polyA_site      3285
                     /note="putative"
ORIGIN

  Query Match             9.0%;  Score 392.8;  DB 4;  Length 3285;
  Best Local Similarity   53.3%;  Pred. No. 5.5e-109;
  Matches 1021;  Conservative    0;  Mismatches  812;  Indels   81;  Gaps    6;

Qy       2353 CAACGTTTGAGAAGCATATTGAGGCCAGACCTTTCAATCTCAAAGCATCGGTGCACATGC 2412

Db        876 CCATCTTAGAACATCAGATTCAAGTCAGACCTTTTAATGCGTTGAAGACAAAGAGTATGA 935

Qy       2413 GTGAACTCAACCCTTCAGATATAGACAAATTGGTTTCTGTTAAAGGAATGGTTATCCGGT 2472

Db        936 GAAACTTGAATCCAGAAGACATTGATCAGCTCATCACCATCAGTGGCATGGTCATCAGAA 995

Qy       2473 GCAGTTCTATCATACCTGAAATTAAGGGGGCCTTCTTCAAATGTTTAGTGTGTGGTCACT 2532

Db        996 CATCACAGCTGATTCCGGAGATGCAGGAGGCCTTTTTCCAATGCCAAGTCTGTGCCCACA 1055

Qy       2533 CGCCTCCGCTAGTTACAGTTGTTAAAGGGCGGGTTGAGGAGCCAACAAGGTGTGAAAAGC 2592

Db       1056 CCACCCGGGTGGAGA---TAGATCGAGGCAGAATTGCTGAGCCCTGCAGTTGTGTGCACT 1112

Qy       2593 CAGAATGTGCAGCACGGAATGCTATGTCTCTTATTCACAATCGATGCACTTTTGCAAATA 2652

Db       1113 GCCACACTACCCACAGC------ATGGCACTGATCCACAACCGATCATTCTTCTCTGACA 1166

Qy       2653 AGCAGATAGTGCGTCTTCAAGAAACTCCAGATGCCATTCCTGAAGGAGAGACTCCACACA 2712

Db       1167 AGCAAATGATCAAACTTCAAGAGTCTCCTGAAGACATGCCTGCTGGGCAGACACCTCACA 1226

Qy       2713 CAGTCAGCATGTGTTTATACAACACTATGGTTGATGCTGTGAAGCCTGGAGATCGTATTG 2772

Db       1227 CTATTGTCCTTTTTGCCCACAATGACCTTGTTGACAAGGTTCAACCAGGGGACAGAGTGA 1286

Qy       2773 AGGTAACAGGAGTTTTCAAGGCCATGGCAGTTCGAGTTGGTCCGAATCAACGAACATTAC 2832

Db       1287 ACGTCACAGGCATATATCGAGCAGTACCAATTCGAGTTAATCCAAGAGTGAGCAACGTGA 1346

Qy       2833 GAGCATTGTATAAGACCTACATCGATTGCGTGCACGTCAAGAAGTCTGACAGGGGTCGAC 2892

Db       1347 AGTCTGTCTATAAAACCCACATTGATGTCATTCATTATCGGAAAACGGATGCAAAACGTC 1406

Qy       2893 TGCAAACTGAAGATCCTATGGAGATGGATAAGGAGAATGATATGTATGCTGGGTATCATG 2952

Db       1407 TGCA---------------------------------------------TGGCCTTGATG 1421

Qy       2953 AAAGTGATACTTCAGAAGCTGCTAATGAAGCAAAGATTCAAAAACTTAAAGAGCTGTCCA 3012

Db       1422 AAGAAGCAGAACAGAAACTTTTTTCAGAGAAACGTGTGAAATTGCTTAAGGAACTTTCCA 1481

Qy       3013 AGCTCCCGGGCATTTATGATAGACTTTCAAGGTCGCTGGCTCCAAGCATTTGGGAGCTTG 3072

Db       1482 GGAAGCCAGATATTTATGAGCGGCTTGCTTCAGCCTTGGCTCCCAGCATTTATGAACATG 1541

Qy       3073 AAGATATTAAAAAGGGTCTTCTTTGCCAGCTCTTTGGTG------GGAAGGCTAAGAAAA 3126

Db       1542 AAGATATCAAAAAGGGAATCTTACTTCAGCTCTTTGGTGGAACAAGGAAGGATTTCAGTC 1601

Qy       3127 TTCCATCTGGAGCATCTTTCCGAGGTGACATCAATGTTTTACTTGTTGGGGACCCTGGTA 3186

Db       1602 ACACTGGGAGGGGTAAATTCCGTGCTGAGATCAACATCCTTCTGTGTGGGGACCCTGGCA 1661

Qy       3187 CCAGTAAATCTCAGCTGCTTCAGTATGTGCACAAGATAGCTCCTCGTGGAATCTACACTA 3246

Db       1662 CCAGCAAGTCCCAGCTGCTACAGTATGTGTACAACCTGGTCCCCAGAGGCCAGTACACGT 1721

Qy       3247 GTGGGCGAGGAAGTTCGGCGGTTGGGCTGACAGCGTATGTAACGAAGGATCCAGAAACTC 3306

Db       1722 CTGGAAAAGGCTCCAGTGCGGTCGGCCTCACCGCCTATGTGATGAAAGACCCTGAGACCA 1781

Qy       3307 GAGAGACGGTATTGGAGAGCGGAGCTTTGGTTCTTAGTGATCGTGGGATATGCTGTATCG 3366
```

```
Db    1782 GGCAGCTTGTCCTCCAGACAGGTGCCCTCGTCCTGAGTGACAATGGGATATGCTGCATCG 1841

Qy    3367 ATGAGTTCGACAAAATGTCTGATAATGCCCGAAGCATGCTTCATGAGGTAATGGAGCAAC 3426
Db    1842 ATAAGTTTGACAAAATGAATGAAAGCACAAGGTCTGTGCTGCATGAGGTCATGGAACAGC 1901

Qy    3427 AAACGGTATCTGTAGCCAAAGGGGGTATCATTGCCTCGCTGAACGCTCGGACGTCTGTCC 3486
Db    1902 AGACTCTGTCCATTGCAAAGGCTGGGATCATCTGTCAGCTCAATGCGCGCACCTCTGTCC 1961

Qy    3487 TTGCATGTGCAAATCCTAGTGGGTCCCGATACAATGCGCGCCTTTCTGTGATTGATAACA 3546
Db    1962 TGGCAGCAGCAAATCCTATTGAGTCTCAGTGGAATCCTAAAAAAACAACCATTGAAAATA 2021

Qy    3547 TCCAGCTTCCTCCAACTCTACTTTCTAGATTTGATTTAATTTACTTAATGCTCGACAAAC 3606
Db    2022 TCCAACTACCGCACACATTGTTGTCAAGGTTTGATCTCATTTTCCTCATGCTAGACCCTC 2081

Qy    3607 CAGACGAGCAAAACGATCGTCGTCTCGCCAGGCATCTCGTGGCTTTACACTATGAAAACT 3666
Db    2082 AGGATGAGGCATATGACCGGCGTCTAGCTCATCACCTGGTTTCATTGTACTACCAAAGTG 2141

Qy    3667 ATGAAGTTTCAAAGCAGGACGCCTTAGATCTACAAACACTTACCGCGTATATCACCTATG 3726
Db    2142 AGGAGCAAGTGGAGGAGGAGTTCCTGGACATGGCCGTGCTGAAAGACTACATTGCATATG 2201

Qy    3727 CTCGTCAGCATGTACATCCTACATTAAGTGATGAAGCTGCTGAAGATTTGATTAATGGCT 3786
Db    2202 CCCATAGTACCATCATGCCCCGACTGAGTGAGGAGGCCAGCCAGGCTCTCATTGAGGCTT 2261

Qy    3787 ATGTTGAGATGCGCCAAAAGGGCAACTTTCCTGGAAGCAGTAAAAAGGTGATAACAGCCA 3846
Db    2262 ATGTAAACATGAGGAAGAT------------TGGGAGTAGCCGGGGGATGGTTTCTGCTT 2309

Qy    3847 CACCTCGGCAACTCGAAAGTATGATTCGTATCAGTGAAGCCCTAGCTCGAATGAGATTTT 3906
Db    2310 ACCCTCGACAGCTAGAGTCATTAATTCGCTTAGCAGAAGCCCATGCTAAAGTAAGATTTT 2369

Qy    3907 CTGAAGTGGTAGAGAAAGTTGATGCAGCAGAAGCTGTGCGCCTTTTAGACGTCGCTTTGC 3966
Db    2370 CAAACAAAGTTGAAGCAATTGATGTGGAAGAGGCAAAACGCCTCCACCGGGAGGCTCTGA 2429

Qy    3967 AGCAATCTGCTACTGATCATGCAACAGGTACGATAGACATGGATCTTATCACGACTGGAG 4026
Db    2430 AGCAGTCTGCAACTGACCCTCGTACTGGCATGGTGGATATTTCTATTCTTACTACAGGAA 2489

Qy    4027 TGTCGGCCAGCGAGCGTATTCGTCGGGCCAACTTGCTAGCTGCTCTGCGAGAGCTTATAG 4086
Db    2490 TGAGTGCCACTTCTCGTAAACGGAAAGAAGAATTAGCTGAAGCATTGAGAAAACTTATTT 2549

Qy    4087 CAGATAAAATTTCACCTGGCAGCTCCTCTGGCTTGAAGACCAGTCAGCTTCTTGAGGATA 4146
Db    2550 TATCTAA---------GGGTAAAACACCAGCCTTAAAGTACCAACAGCTGTTTGAGGATA 2600

Qy    4147 TCCGGAGCCAAAGCAGTGTGGACGTTAGTTTGCAGGATATTAAAAATGCTCTGGGTAGCC 4206
Db    2601 TTCGGGGACAGTCTGACACAGCAATTACCAAGGACATGTTTGAAGAAGCCCTGCGAGCTT 2660

Qy    4207 TCCAAGGAGAAGGCTTTCTTACTGTCCATGGTGACATAGTCAAGAGAGTTTGAG 4260
Db    2661 TGGCTGATGATGATTTCCTAACAGTGACTGGGAAGACTGTCCGCCTGCTGTGAG 2714
```

```
RESULT 15
AK088796
LOCUS       AK088796             3301 bp    mRNA    linear   HTC 03-APR-2004
DEFINITION  Mus musculus 2 days neonate thymus thymic cells cDNA, RIKEN
            full-length enriched library, clone:E430026B21 product:mini
            chromosome maintenance deficient 4 homolog (S. cerevisiae), full
            insert sequence.
ACCESSION   AK088796
VERSION     AK088796.1  GI:26353895
KEYWORDS    HTC; CAP trapper.
SOURCE      Mus musculus (house mouse)
  ORGANISM  Mus musculus
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Euarchontoglires; Glires; Rodentia;
            Sciurognathi; Muridae; Murinae; Mus.
REFERENCE   1
  AUTHORS   Carninci,P. and Hayashizaki,Y.
  TITLE     High-efficiency full-length cDNA cloning
  JOURNAL   Meth. Enzymol. 303, 19-44 (1999)
   PUBMED   10349636
REFERENCE   2
  AUTHORS   Carninci,P., Shibata,Y., Hayatsu,N., Sugahara,Y., Shibata,K.,
            Itoh,M., Konno,H., Okazaki,Y., Muramatsu,M. and Hayashizaki,Y.
  TITLE     Normalization and subtraction of cap-trapper-selected cDNAs to
            prepare full-length cDNA libraries for rapid discovery of new genes
  JOURNAL   Genome Res. 10 (10), 1617-1630 (2000)
   PUBMED   11042159
REFERENCE   3
  AUTHORS   Shibata,K., Itoh,M., Aizawa,K., Nagaoka,S., Sasaki,N., Carninci,P.,
            Konno,H., Akiyama,J., Nishi,K., Kitsunai,T., Tashiro,H., Itoh,M.,
            Sumi,N., Ishii,Y., Nakamura,S., Hazama,M., Nishine,T., Harada,A.,
            Yamamoto,R., Matsumoto,H., Sakaguchi,S., Ikegami,T., Kashiwagi,K.,
            Fujiwake,S., Inoue,K., Togawa,Y., Izawa,M., Ohara,E., Watahiki,M.,
            Yoneda,Y., Ishikawa,T., Ozawa,K., Tanaka,T., Matsuura,S., Kawai,J.,
            Okazaki,Y., Muramatsu,M., Inoue,Y., Kira,A. and Hayashizaki,Y.
  TITLE     RIKEN integrated sequence analysis (RISA) system--384-format
            sequencing pipeline with 384 multicapillary sequencer
  JOURNAL   Genome Res. 10 (11), 1757-1771 (2000)
   PUBMED   11076861
REFERENCE   4
  AUTHORS   The RIKEN Genome Exploration Research Group Phase II Team and the
            FANTOM Consortium.
  TITLE     Functional annotation of a full-length mouse cDNA collection
  JOURNAL   Nature 409, 685-690 (2001)
REFERENCE   5
  AUTHORS   The FANTOM Consortium and the RIKEN Genome Exploration Research
            Group Phase I & II Team.
  TITLE     Analysis of the mouse transcriptome based on functional annotation
            of 60,770 full-length cDNAs
  JOURNAL   Nature 420, 563-573 (2002)
REFERENCE   6  (bases 1 to 3301)
  AUTHORS   Adachi,J., Aizawa,K., Akimura,T., Arakawa,T., Bono,H., Carninci,P.,
            Fukuda,S., Furuno,M., Hanagaki,T., Hara,A., Hashizume,W.,
            Hayashida,K., Hayatsu,N., Hiramoto,K., Hiraoka,T., Hirozane,T.,
            Hori,F., Imotani,K., Ishii,Y., Itoh,M., Kagawa,I., Kasukawa,T.,
            Katoh,H., Kawai,J., Kojima,Y., Kondo,S., Konno,H., Kouda,M.,
            Koya,S., Kurihara,C., Matsuyama,T., Miyazaki,A., Murata,M.,
            Nakamura,M., Nishi,K., Nomura,K., Numazaki,R., Ohno,M., Ohsato,N.,
            Okazaki,Y., Saito,R., Saitoh,H., Sakai,C., Sakai,K., Sakazume,N.,
            Sano,H., Sasaki,D., Shibata,K., Shinagawa,A., Shiraki,T.,
            Sogabe,Y., Tagami,M., Tagawa,A., Takahashi,F., Takaku-Akahira,S.,
            Takeda,Y., Tanaka,T., Tomaru,A., Toya,T., Yasunishi,A.,
            Muramatsu,M. and Hayashizaki,Y.
  TITLE     Direct Submission
  JOURNAL   Submitted (16-APR-2002) Yoshihide Hayashizaki, The Institute of
            Physical and Chemical Research (RIKEN), Laboratory for Genome
            Exploration Research Group, RIKEN Genomic Sciences Center (GSC),
            RIKEN Yokohama Institute; 1-7-22 Suehiro-cho, Tsurumi-ku, Yokohama,
            Kanagawa 230-0045, Japan (E-mail:genome-res@gsc.riken.jp,
            URL:http://genome.gsc.riken.jp/, Tel:81-45-503-9222,
            Fax:81-45-503-9216)
COMMENT     cDNA library was prepared and sequenced in Mouse Genome
            Encyclopedia Project of Genome Exploration Research Group in Riken
            Genomic Sciences Center and Genome Science Laboratory in RIKEN.
            Division of Experimental Animal Research in Riken contributed to
            prepare mouse tissues.
            Tissues were provided by Dr. John Todd (Dept. of Medical Genetics
            Wellcome Trust Centre for Molecular Mechanisms in Disease Wellcome
            Trust/MRC building Addenbrookes Hospital Cambridge) whose
            assistance we gratefully acknowledge.
            Please visit our web site for further details.
            URL:http://genome.gsc.riken.jp/
            URL:http://fantom.gsc.riken.jp/.
```

```
FEATURES             Location/Qualifiers
     source          1..3301
                     /organism="Mus musculus"
                     /mol_type="mRNA"
                     /strain="NOD"
                     /db_xref="FANTOM_DB:E430026B21"
                     /db_xref="taxon:10090"
                     /clone="E430026B21"
                     /cell_type="thymic cells"
                     /tissue_type="thymus"
                     /clone_lib="RIKEN full-length enriched mouse cDNA library"
                     /dev_stage="2 days neonate"
     CDS             141..2729
                     /note="unnamed protein product; mini chromosome
                     maintenance deficient 4 homolog (S. cerevisiae)
                     (MGD|MGI:103199, GB|NM_008565, evidence: BLASTN, 100%,
                     match=3221)
                     putative"
                     /codon_start=1
                     /protein_id="BAC40578.1"
                     /db_xref="GI:26353896"
                     /translation="MSSPASTPSRRSSRRGRVTPTQSLRSEESRSSPNRRRRGEDSST
                     GELLPMPTSPGADLQSPPAQNALFSSPPQMHSLAIPLDFDVSSPLTYGTPSSRVEGTP
                     RSGVRGTPVRQRPDLGSARKGLQVDLQSDGAAAEDIVPSEQSLGQKLVIWGTDVNVAT
                     CKENFQRFLQCFTDPLAKEEENVGIDITQPLYMQQLGEINITGEPFLNVNCEHIKSFS
                     KNLYRQLISYPQEVIPTFDMAVNEIFFDRYPDSILEHQIQVRPFNALKTKSMRNLNPE
                     DIDQLITISGMVIRTSQLIPEMQEAFFQCQVCAHTTRVEIDRGRIAEPCSCVHCHTTH
                     SMALIHNRSFFSDKQMIKLQESPEDMPAGQTPHTIVLFAHNDLVDKVQPGDRVNVTGI
                     YRAVPIRVNPRVSNVKSVYKTHIDVIHYRKTDAKRLHGLDEEAEQKLFSEKRVKLLKE
                     LSRKPDIYERLASALAPSIYEHEDIKKGILLQLFGGTRKDFSHTGRGKFRAEINILLC
                     GDPGTSKSQLLQYVYNLVPRGQYTSGKGSSAVGLTAYVMKDPETRQLVLQTGALVLSD
                     NGICCIDEFDKMNESTRSVLHEVMEQQTLSIAKAGIICQLNARTSVLAAANPIESQWN
                     PKKTTIENIQLPHTLLSRFDLIFLMLDPQDEAYDRRLAHHLVSLYYQSEEQVEEEFLD
                     MAVLKDYIAYAHSTIMPRLSEEASQALIEAYVNMRKIGSSRGMVSAYPRQLESLIRLA
                     EAHAKVRFSNKVEAIDVEEAKRLHREALKQSATDPRTGIVDISILTTGMSATSRKRKE
                     ELAEALRKLILSKGKTPALKYQQLFEDIRGQSDTAITKDMFEEALRALADDDFLTVTG
                     KTVRLL"
     polyA_signal    3284..3289
                     /note="putative"
     polyA_site      3301
                     /note="putative"
ORIGIN

  Query Match             9.0%;  Score 392.8;  DB 4;  Length 3301;
  Best Local Similarity   53.3%;  Pred. No. 5.5e-109;
  Matches 1021;  Conservative    0;  Mismatches  812;  Indels   81;  Gaps    6;

Qy       2353 CAACGTTTGAGAAGCATATTGAGGCCAGACCTTTCAATCTCAAAGCATCGGTGCACATGC 2412
Db        892 CCATCTTAGAACATCAGATTCAAGTCAGACCTTTTAATGCGTTGAAGACAAAGAGTATGA 951
Qy       2413 GTGAACTCAACCCTTCAGATATAGACAAATTGGTTTCTGTTAAAGGAATGGTTATCCGGT 2472
Db        952 GAAACTTGAATCCAGAAGACATTGATCAGCTCATCACCATCAGTGGCATGGTCATCAGAA 1011
Qy       2473 GCAGTTCTATCATACCTGAAATTAAGGGGGCCTTCTTCAAATGTTTAGTGTGTGGTCACT 2532
Db       1012 CATCACAGCTGATTCCGGAGATGCAGGAGGCCTTTTTCCAATGCCAAGTCTGTGCCCACA 1071
Qy       2533 CGCCTCCGCTAGTTACAGTTGTTAAAGGGCGGGTTGAGGAGCCAACAAGGTGTGAAAAGC 2592
Db       1072 CCACCCGGGTGGAGA---TAGATCGAGGCAGAATTGCTGAGCCCTGCAGTTGTGTGCACT 1128
Qy       2593 CAGAATGTGCAGCACGGAATGCTATGTCTCTTATTCACAATCGATGCACTTTTGCAAATA 2652
Db       1129 GCCACACTACCCACAGC------ATGGCACTGATCCACAACCGATCATTCTTCTCTGACA 1182
Qy       2653 AGCAGATAGTGCGTCTTCAAGAAACTCCAGATGCCATTCCTGAAGGAGAGACTCCACACA 2712
Db       1183 AGCAAATGATCAAACTTCAAGAGTCTCCTGAAGACATGCCTGCTGGGCAGACACCTCACA 1242
Qy       2713 CAGTCAGCATGTGTTTATACAACACTATGGTTGATGCTGTGAAGCCTGGAGATCGTATTG 2772
Db       1243 CTATTGTCCTTTTTGCCCACAATGACCTTGTTGACAAGGTTCAACCAGGGGACAGAGTGA 1302
Qy       2773 AGGTAACAGGAGTTTTCAAGGCCATGGCAGTTCGAGTTGGTCCGAATCAACGAACATTAC 2832
Db       1303 ACGTCACAGGCATATATCGAGCAGTACCAATTCGAGTTAATCCAAGAGTGAGCAACGTGA 1362
Qy       2833 GAGCATTGTATAAGACCTACATCGATTGCGTGCACGTCAAGAAGTCTGACAGGGGTCGAC 2892
Db       1363 AGTCTGTCTATAAAACCCACATTGATGTCATTCATTATCGGAAAACGGATGCAAAACGTC 1422
Qy       2893 TGCAAACTGAAGATCCTATGGAGATGGATAAGGAGAATGATATGTATGCTGGGTATCATG 2952
Db       1423 TGCA-----------------------------------------TGGCCTTGATG 1437
Qy       2953 AAAGTGATACTTCAGAAGCTGCTAATGAAGCAAAGATTCAAAAACTTAAAGAGCTGTCCA 3012
Db       1438 AAGAAGCAGAACAGAAACTTTTTTCAGAGAAACGTGTGAAATTGCTTAAGGAACTTTCCA 1497
Qy       3013 AGCTCCCGGGCATTTATGATAGACTTTCAAGGTCGCTGGCTCCAAGCATTTGGGAGCTTG 3072
Db       1498 GGAAGCCAGATATTTATGAGCGGCTTGCTTCAGCCTTGGCTCCCAGCATTTATGAACATG 1557
Qy       3073 AAGATATTAAAAAGGGTCTTCTTTGCCAGCTCTTTGGTG------GGAAGGCTAAGAAAA 3126
Db       1558 AAGATATCAAAAAGGGAATCTTACTTCAGCTCTTTGGTGGAACAAGGAAGGATTTCAGTC 1617
Qy       3127 TTCCATCTGGAGCATCTTTCCGAGGTGACATCAATGTTTTACTTGTTGGGGACCCTGGTA 3186
Db       1618 ACACTGGGAGGGGTAAATTCCGTGCTGAGATCAACATCCTTCTGTGTGGGGACCCTGGCA 1677
Qy       3187 CCAGTAAATCTCAGCTGCTTCAGTATGTGCACAAGATAGCTCCTCGTGGAATCTACACTA 3246
Db       1678 CCAGCAAGTCCCAGCTGCTACAGTATGTGTACAACCTGGTCCCCAGAGGCCAGTACACGT 1737
Qy       3247 GTGGGCGAGGAAGTTCGGCGGTTGGGCTGACAGCGTATGTAACGAAGGATCCAGAAACTC 3306
Db       1738 CTGGAAAAGGCTCCAGTGCGGTCGGCCTCACCGCCTATGTGATGAAAGACCCTGAGACCA 1797
Qy       3307 GAGAGACGGTATTGGAGAGCGGAGCTTTGGTTCTTAGTGATCGTGGGATATGCTGTATCG 3366
Db       1798 GGCAGCTTGTCCTCCAGACAGGTGCCCTCGTCCTGAGTGACAATGGGATATGCTGCATCG 1857
Qy       3367 ATGAGTTCGACAAAATGTCTGATAATGCCCGAAGCATGCTTCATGAGGTAATGGAGCAAC 3426
Db       1858 ATGAGTTTGACAAAATGAATGAAAGCACAAGGTCTGTGCTGCATGAGGTCATGGAACAGC 1917
Qy       3427 AAACGGTATCTGTAGCCAAAGGGGGTATCATTGCCTCGCTGAACGCTCGGACGTCTGTCC 3486
Db       1918 AGACTCTGTCCATTGCAAAGGCTGGGATCATCTGTCAGCTCAATGCGCGCACCTCTGTCC 1977
Qy       3487 TTGCATGTGCAAATCCTAGTGGGTCCCGATACAATGCGCGCCTTTCTGTGATTGATAACA 3546
Db       1978 TGGCAGCAGCAAATCCTATTGAGTCTCAGTGGAATCCTAAAAAAACAACCATTGAAAATA 2037
Qy       3547 TCCAGCTTCCTCCAACTCTACTTTCTAGATTTGATTTAATTTACTTAATGCTCGACAAAC 3606
Db       2038 TCCAACTACCGCACACATTGTTGTCAAGGTTTGATCTCATTTTCCTCATGCTAGACCCTC 2097
Qy       3607 CAGACGAGCAAAACGATCGTCGTCTCGCCAGGCATCTCGTGGCTTTACACTATGAAAACT 3666
Db       2098 AGGATGAGGCATATGACCGGCGTCTAGCTCATCACCTGGTTTCATTGTACTACCAAAGTG 2157
Qy       3667 ATGAAGTTTCAAAGCAGGACGCCTTAGATCTACAAACACTTACCGCGTATATCACCTATG 3726
Db       2158 AGGAGCAAGTGGAGGAGGAGTTCCTGGACATGGCCGTGCTGAAAGACTACATTGCATATG 2217
Qy       3727 CTCGTCAGCATGTACATCCTACATTAAGTGATGAAGCTGCTGAAGATTTGATTAATGGCT 3786
Db       2218 CCCATAGTACCATCATGCCCCGACTGAGTGAGGAGGCCAGCCAGGCTCTCATTGAGGCTT 2277
Qy       3787 ATGTTGAGATGCGCCAAAAGGGCAACTTTCCTGGAAGCAGTAAAAAGGTGATAACAGCCA 3846
Db       2278 ATGTAAACATGAGGAAGAT-----------TGGGAGTAGCCGGGGGATGGTTTCTGCTT 2325
```

```
Qy   3847 CACCTCGGCAACTCGAAAGTATGATTCGTATCAGTGAAGCCCTAGCTCGAATGAGATTTT 3906
Db   2326 ACCCTCGACAGCTAGAGTCATTAATTCGCTTAGCAGAAGCCCATGCTAAAGTAAGATTTT 2385

Qy   3907 CTGAAGTGGTAGAGAAAGTTGATGCAGCAGAAGCTGTGCGCCTTTTAGACGTCGCTTTGC 3966
Db   2386 CAAACAAAGTTGAAGCAATTGATGTGGAAGAGGCAAAACGCCTCCACCGGGAGGCTCTGA 2445

Qy   3967 AGCAATCTGCTACTGATCATGCAACAGGTACGATAGACATGGATCTTATCACGACTGGAG 4026
Db   2446 AGCAGTCTGCAACTGACCCTCGTACTGGCATTGTGGATATTTCTATTCTTACTACAGGAA 2505

Qy   4027 TGTCGGCCAGCGAGCGTATTCGTCGGGCCAACTTGCTAGCTGCTCTGCGAGAGCTTATAG 4086
Db   2506 TGAGTGCCACTTCTCGTAAACGGAAAGAAGAATTAGCTGAAGCATTGAGAAAACTTATTT 2565

Qy   4087 CAGATAAAATTTCACCTGGCAGCTCCTCTGGCTTGAAGACCAGTCAGCTTCTTGAGGATA 4146
Db   2566 TATCTAA---------GGGTAAAACACCAGCCTTAAAGTACCAACAGCTGTTTGAGGATA 2616

Qy   4147 TCCGGAGCCAAAGCAGTGTGGACGTTAGTTTGCAGGATATTAAAAATGCTCTGGGTAGCC 4206
Db   2617 TTCGGGGACAGTCTGACACAGCAATTACCAAGGACATGTTTGAAGAAGCCCTGCGAGCTT 2676

Qy   4207 TCCAAGGAGAAGGCTTTCTTACTGTCCATGGTGACATAGTCAAGAGAGTTTGAG 4260
Db   2677 TGGCTGATGATGATTTCCTAACAGTGACTGGGAAGACTGTCCGCCTGCTCTGAG 2730
```

Search completed: December 6, 2005, 09:08:19
Job time : 10586 secs

OM protein - protein search, using sw model

Run on: December 7, 2005, 17:03:05 ; Search time 191 Seconds
(without alignments)
2072.672 Million cell updates/sec

Title: US-10-768-511-8
Perfect score: 4609
Sequence: 1 MENNDALDIGAVSSPYPSQS..........LGSLQGEGFLTVHGDIVKRV 901

Scoring table: BLOSUM62
Gapop 10.0 , Gapext 0.5

Searched: 2443163 seqs, 439378781 residues

Total number of hits satisfying chosen parameters: 2443163

Minimum DB seq length: 0
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries

Database : A_Geneseq_21:*
1: geneseqp1980s:*
2: geneseqp1990s:*
3: geneseqp2000s:*
4: geneseqp2001s:*
5: geneseqp2002s:*
6: geneseqp2003as:*
7: geneseqp2003bs:*
8: geneseqp2004s:*
9: geneseqp2005s:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	% Query Match	Length	DB	ID	Description
1	4609	100.0	901	5	AAG66003	Aag66003 P. patens
2	4609	100.0	901	7	AAE39980	Aae39980 Physcomit
3	2550	55.3	881	8	ADY09589	Ady09589 Plant ful
4	1883	40.9	1023	6	ABJ26500	Abj26500 Aspergill
5	1860	40.4	874	6	ABJ25900	Abj25900 Aspergill
6	1835.5	39.8	989	7	ADB70288	Adb70288 C. neofor
7	1835	39.8	960	8	ADN21148	Adn21148 Bacterial
8	1833	39.8	933	8	ADS44081	Ads44081 Bacterial
9	1817.5	39.4	923	7	ADD18710	Add18710 Human dis
10	1817.5	39.4	923	8	ADO19848	Ado19848 Human PRO
11	1817.5	39.4	924	8	ADU06294	Adu06294 Novel bro
12	1814.5	39.4	863	8	ADR69821	Adr69821 Human min
13	1810.5	39.3	863	4	AAB95203	Aab95203 Human pro
14	1810.5	39.3	863	8	ABM80987	Abm80987 Tumour-as
15	1754	38.1	910	5	ABP73643	Abp73643 Candida a
16	1732.5	37.6	918	8	ADP83536	Adp83536 Breast sp
17	1686.5	36.6	866	4	ABB58299	Abb58299 Drosophil
18	1670.5	36.2	456	8	ADY13148	Ady13148 Plant ful
19	1647	35.7	820	8	ABM82800	Abm82800 Human dia
20	1500	32.5	712	8	ADQ19271	Adq19271 Human sof
21	1375	29.8	724	3	AAB56475	Aab56475 Human pro
22	1199.5	26.0	656	8	ADP83535	Adp83535 Breast sp
23	1071	23.2	682	4	AAB62036	Aab62036 P. furios
24	1020.5	22.1	666	8	ADS43108	Ads43108 Bacterial
25	1016.5	22.1	760	8	ADS44134	Ads44134 Bacterial
26	1003	21.8	814	6	ABJ26042	Abj26042 Aspergill
27	997.5	21.6	287	8	ADX68835	Adx68835 Plant ful
28	985.5	21.4	718	6	ABJ25741	Abj25741 Aspergill
29	983.5	21.3	713	6	ABJ26341	Abj26341 Aspergill
30	983.5	21.3	809	5	ABP73372	Abp73372 Candida a
31	976.5	21.2	720	4	ABB61153	Abb61153 Drosophil
32	973.5	21.1	632	6	ABJ25442	Abj25442 Aspergill
33	973	21.1	699	8	ADS44402	Ads44402 Bacterial
34	972	21.1	743	8	ADX95960	Adx95960 Plant ful
35	970	21.0	766	3	AAB43937	Aab43937 Human can
36	963.5	20.9	720	5	AAO22751	Aao22751 Maize pro
37	963	20.9	775	6	ABR53639	Abr53639 Protein s
38	963	20.9	775	7	ADK62368	Adk62368 Disease t
39	963	20.9	775	8	ADN19147	Adn19147 Bacterial
40	961.5	20.9	719	8	ADO19649	Ado19649 Human PRO
41	961.5	20.9	719	8	ADO19790	Ado19790 Human PRO
42	961.5	20.9	719	8	ADU06531	Adu06531 Novel bro
43	961.5	20.9	719	9	ADY16524	Ady16524 PRO polyp
44	961.5	20.9	881	8	ADN22948	Adn22948 Bacterial
45	960.5	20.8	719	8	ADQ09276	Adq09276 Human MCM

ALIGNMENTS

RESULT 1
AAG66003
ID AAG66003 standard; protein; 901 AA.
XX
AC AAG66003;
XX
DT 27-FEB-2002 (first entry)
XX
DE P. patens cell cycle protein 2 (CC-2).
XX
KW Cell Cycle Stress-Related Protein; CCSRP; cell cycle protein; CC-1; CC-2;
KW CC-3; environmental stress.
XX
OS Physcomitrella patens.
XX
PN WO200177354-A2.
XX
PD 18-OCT-2001.
XX
PF 06-APR-2001; 2001WO-US011294.
XX
PR 07-APR-2000; 2000US-0196001P.
XX
PA (BADI) BASF PLANT SCI GMBH.
XX
PI Costa E SilvaO, Bohnert HJ, Van Thielen N, Chen R;
PI Sarria-Millan R;
XX
DR WPI; 2002-049151/06.
DR N-PSDB; AAI67613.
XX
PT Novel Cell Cycle Stress-Related Protein useful for increasing tolerance
PT to environmental stress, is selected from Cell Cycle Proteins 1-3, or
PT their orthologs.
XX
PS Claim 3; Fig 3B; 90pp; English.
XX
CC The invention relates to a Cell Cycle Stress-Related Protein (CCSRP),
CC isolated from Physcomitrella patens, and selected from Cell Cycle (CC)-1
CC protein, CC-2 protein, CC-3 protein, or their orthologs. The CCSRP and
CC encoding nucleic acids are useful for increasing tolerance to
CC environmental stress selected from salinity, drought and temperature, in
CC transgenic plants including monocot and dicot selected from maize, wheat,
CC rye, oat, triticale, rice, barley, soybean, peanut, cotton, rapeseed,
CC canola, manihot, pepper, sunflower, tagetes, solanaceous plants, potato,
CC tobacco, eggplant, tomato, Vicia species, pea, alfalfa, coffee, cacao,
CC tea, Salix species, oil palm, coconut, perennial grass and forage crops.

```
CC   The nucleic acid is also useful for identifying organisms e.g. P. patens
CC   in a mixed population of microorganisms. The nucleic acids are also
CC   useful for evolutionary and protein structural studies. The proteins and
CC   nucleic acids are useful as markers for specific regions of the genome.
CC   The present sequence represents the P. patens CC-21 protein
XX
SQ   Sequence 901 AA;

  Query Match             100.0%;  Score 4609;  DB 5;  Length 901;
  Best Local Similarity   100.0%;  Pred. No. 0;
  Matches  901;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy          1 MENNDALDIGAVSSPYPSQSEGVSTPLPQVTSPSFDNAASPVAGRRAVRQTPTSAVRRRG 60
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db          1 MENNDALDIGAVSSPYPSQSEGVSTPLPQVTSPSFDNAASPVAGRRAVRQTPTSAVRRRG 60

Qy         61 RETDSARRRRSRSRSLGNSVYSSPYDAGTPGTPGTPVATPVYATPVGTPMGTPSFHRGTP 120
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db         61 RETDSARRRRSRSRSLGNSVYSSPYDAGTPGTPGTPVATPVYATPVGTPMGTPSFHRGTP 120

Qy        121 QYKQRSELGSQGKPLHRRRRSQSREPGHRSPSREPSADGRPSESAEPDDTLGGEYAYVWG 180
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        121 QYKQRSELGSQGKPLHRRRRSQSREPGHRSPSREPSADGRPSESAEPDDTLGGEYAYVWG 180

Qy        181 TNVNIPDVLRAIRRFLHNYRSSAHDLNSKYIQIIEETVEREEDTLNIDMSDIYDHDPDLY 240
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        181 TNVNIPDVLRAIRRFLHNYRSSAHDLNSKYIQIIEETVEREEDTLNIDMSDIYDHDPDLY 240

Qy        241 AKIVRYPLDIIPLLDTECQEVATSLLPTFEKHIEARPFNLKASVHMRELNPSDIDKLVSV 300
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        241 AKIVRYPLDIIPLLDTECQEVATSLLPTFEKHIEARPFNLKASVHMRELNPSDIDKLVSV 300

Qy        301 KGMVIRCSSIIPEIKGAFFKCLVCGHSPPLVTVVKGRVEEPTRCEKPECAARNAMSLIHN 360
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        301 KGMVIRCSSIIPEIKGAFFKCLVCGHSPPLVTVVKGRVEEPTRCEKPECAARNAMSLIHN 360

Qy        361 RCTFANKQIVRLQETPDAIPEGETPHTVSMCLYNTMVDAVKPGDRIEVTGVFKAMAVRVG 420
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        361 RCTFANKQIVRLQETPDAIPEGETPHTVSMCLYNTMVDAVKPGDRIEVTGVFKAMAVRVG 420

Qy        421 PNQRTLRALYKTYIDCVHVKKSDRGRLQTEDPMEMDKENDMYAGYHESDTSEAANEAKIQ 480
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        421 PNQRTLRALYKTYIDCVHVKKSDRGRLQTEDPMEMDKENDMYAGYHESDTSEAANEAKIQ 480

Qy        481 KLKELSKLPGIYDRLSRSLAPSIWELEDIKKGLLCQLFGGKAKKIPSGASFRGDINVLLV 540
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        481 KLKELSKLPGIYDRLSRSLAPSIWELEDIKKGLLCQLFGGKAKKIPSGASFRGDINVLLV 540

Qy        541 GDPGTSKSQLLQYVHKIAPRGIYTSGRGSSAVGLTAYVTKDPETRETVLESGALVLSDRG 600
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        541 GDPGTSKSQLLQYVHKIAPRGIYTSGRGSSAVGLTAYVTKDPETRETVLESGALVLSDRG 600

Qy        601 ICCIDEFDKMSDNARSMLHEVMEQQTVSVAKGGIIASLNARTSVLACANPSGSRYNARLS 660
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        601 ICCIDEFDKMSDNARSMLHEVMEQQTVSVAKGGIIASLNARTSVLACANPSGSRYNARLS 660

Qy        661 VIDNIQLPPTLLSRFDLIYLMLDKPDEQNDRRLARHLVALHYENYEVSKQDALDLQTLTA 720
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        661 VIDNIQLPPTLLSRFDLIYLMLDKPDEQNDRRLARHLVALHYENYEVSKQDALDLQTLTA 720

Qy        721 YITYARQHVHPTLSDEAAEDLINGYVEMRQKGNFPGSSKKVITATPRQLESMIRISEALA 780
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        721 YITYARQHVHPTLSDEAAEDLINGYVEMRQKGNFPGSSKKVITATPRQLESMIRISEALA 780

Qy        781 RMRFSEVVEKVDAAEAVRLLDVALQQSATDHATGTIDMDLITTGVSASERIRRANLLAAL 840
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        781 RMRFSEVVEKVDAAEAVRLLDVALQQSATDHATGTIDMDLITTGVSASERIRRANLLAAL 840

Qy        841 RELIADKISPGSSSGLKTSQLLEDIRSQSSVDVSLQDIKNALGSLQGEGFLTVHGDIVKR 900
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        841 RELIADKISPGSSSGLKTSQLLEDIRSQSSVDVSLQDIKNALGSLQGEGFLTVHGDIVKR 900

Qy        901 V 901
              |
Db        901 V 901


RESULT 2
AAE39980
ID   AAE39980 standard; protein; 901 AA.
XX
AC   AAE39980;
XX
DT   15-JAN-2004  (first entry)
XX
DE   Physcomitrella patens cell cycle protein (CC-2).
XX
KW   Cell cycle stress-related protein; CCSRP; metal; cell cycle protein; CC;
KW   transgenic plant; environmental stress; stress tolerance; salinity;
KW   drought; temperature; chemical; oxidative stress.
XX
OS   Physcomitrella patens.
XX
FH   Key             Location/Qualifiers
FT   Misc-difference 420. .421
FT                   /note= "Encoded by bases from 1406 to 2813"
XX
PN   US2003097675-A1.
XX
PD   22-MAY-2003.
XX
PF   06-APR-2001; 2001US-00828062.
XX
PR   07-APR-2000; 2000US-0196001P.
XX
PA   (SILV/) COSTA E SILVA O.
PA   (BOHN/) BOHNERT H J.
PA   (THIE/) THIELEN N V.
PA   (CHEN/) CHEN R.
PA   (SARR/) SARRIA-MILLAN R.
XX
PI   Costa E SilvaO,  Bohnert HJ,  Thielen NV,  Chen R,  Sarria-Millan R;
XX
DR   WPI; 2003-765533/72.
DR   N-PSDB; AAD60813.
XX
PT   Novel cell cycle stress-related protein capable of conferring stress
PT   tolerance such as tolerance towards salinity, drought, temperature,
PT   chemical, pathogens, to plants upon over-expression.
XX
PS   Claim 14; Fig 3A; 62pp; English.
XX
CC   The invention relates to an isolated cell cycle stress-related protein
CC   (CCSRP) which is chosen from a cell cycle-1 protein (CC-1), CC-2, CC-3
CC   and its orthologues. Recombinant expression vector is useful for
CC   producing a transgenic plant containing a CCSRP coding nucleic acid,
CC   where expression of the nucleic acid in the plant results in increased
CC   tolerance to environmental stress as compared to a wild type variety of
CC   the plant which involves transforming a plant cell with the recombinant
CC   expression vector, generating from the plant cell a transgenic plant with
CC   an increased tolerance to environmental stress as compared to a wild type
CC   variety of the plant. CCSRP is useful for conferring stress tolerance
CC   such as tolerance towards salinity, drought, temperature, metal,
CC   chemical, pathogens and oxidative stress or their combinations to plants.
CC   The present sequence is Physcomitrella patens CC-2 protein
XX
SQ   Sequence 901 AA;

  Query Match             100.0%;  Score 4609;  DB 7;  Length 901;
  Best Local Similarity   100.0%;  Pred. No. 0;
  Matches  901;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy          1 MENNDALDIGAVSSPYPSQSEGVSTPLPQVTSPSFDNAASPVAGRRAVRQTPTSAVRRRG 60
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db          1 MENNDALDIGAVSSPYPSQSEGVSTPLPQVTSPSFDNAASPVAGRRAVRQTPTSAVRRRG 60
```

```
Qy         61 RETDSARRRRSRSRSLGNSVYSSPYDAGTPGTPGTPVATPVYATPVGTPMGTPSFHRGTP 120
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db         61 RETDSARRRRSRSRSLGNSVYSSPYDAGTPGTPGTPVATPVYATPVGTPMGTPSFHRGTP 120

Qy        121 QYKQRSELGSQGKPLHRRRRSQSREPGHRSPSREPSADGRPSESAEPDDTLGGEYAYVWG 180
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        121 QYKQRSELGSQGKPLHRRRRSQSREPGHRSPSREPSADGRPSESAEPDDTLGGEYAYVWG 180

Qy        181 TNVNIPDVLRAIRRFLHNYRSSAHDLNSKYIQIIEETVEREEDTLNIDMSDIYDHDPDLY 240
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        181 TNVNIPDVLRAIRRFLHNYRSSAHDLNSKYIQIIEETVEREEDTLNIDMSDIYDHDPDLY 240

Qy        241 AKIVRYPLDIIPLLDTECQEVATSLLPTFEKHIEARPFNLKASVHMRELNPSDIDKLVSV 300
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        241 AKIVRYPLDIIPLLDTECQEVATSLLPTFEKHIEARPFNLKASVHMRELNPSDIDKLVSV 300

Qy        301 KGMVIRCSSIIPEIKGAFFKCLVCGHSPPLVTVVKGRVEEPTRCEKPECAARNAMSLIHN 360
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        301 KGMVIRCSSIIPEIKGAFFKCLVCGHSPPLVTVVKGRVEEPTRCEKPECAARNAMSLIHN 360

Qy        361 RCTFANKQIVRLQETPDAIPEGETPHTVSMCLYNTMVDAVKPGDRIEVTGVFKAMAVRVG 420
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        361 RCTFANKQIVRLQETPDAIPEGETPHTVSMCLYNTMVDAVKPGDRIEVTGVFKAMAVRVG 420

Qy        421 PNQRTLRALYKTYIDCVHVKKSDRGRLQTEDPMEMDKENDMYAGYHESDTSEAANEAKIQ 480
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        421 PNQRTLRALYKTYIDCVHVKKSDRGRLQTEDPMEMDKENDMYAGYHESDTSEAANEAKIQ 480

Qy        481 KLKELSKLPGIYDRLSRSLAPSIWELEDIKKGLLCQLFGGKAKKIPSGASFRGDINVLLV 540
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        481 KLKELSKLPGIYDRLSRSLAPSIWELEDIKKGLLCQLFGGKAKKIPSGASFRGDINVLLV 540

Qy        541 GDPGTSKSQLLQYVHKIAPRGIYTSGRGSSAVGLTAYVTKDPETRETVLESGALVLSDRG 600
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        541 GDPGTSKSQLLQYVHKIAPRGIYTSGRGSSAVGLTAYVTKDPETRETVLESGALVLSDRG 600

Qy        601 ICCIDEFDKMSDNARSMLHEVMEQQTVSVAKGGIIASLNARTSVLACANPSGSRYNARLS 660
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        601 ICCIDEFDKMSDNARSMLHEVMEQQTVSVAKGGIIASLNARTSVLACANPSGSRYNARLS 660

Qy        661 VIDNIQLPPTLLSRFDLIYLMLDKPDEQNDRRLARHLVALHYENYEVSKQDALDLQTLTA 720
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        661 VIDNIQLPPTLLSRFDLIYLMLDKPDEQNDRRLARHLVALHYENYEVSKQDALDLQTLTA 720

Qy        721 YITYARQHVHPTLSDEAAEDLINGYVEMRQKGNFPGSSKKVITATPRQLESMIRISEALA 780
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        721 YITYARQHVHPTLSDEAAEDLINGYVEMRQKGNFPGSSKKVITATPRQLESMIRISEALA 780

Qy        781 RMRFSEVVEKVDAAEAVRLLDVALQQSATDHATGTIDMDLITTGVSASERIRRANLLAAL 840
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        781 RMRFSEVVEKVDAAEAVRLLDVALQQSATDHATGTIDMDLITTGVSASERIRRANLLAAL 840

Qy        841 RELIADKISPGSSSGLKTSQLLEDIRSQSSVDVSLQDIKNALGSLQGEGFLTVHGDIVKR 900
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        841 RELIADKISPGSSSGLKTSQLLEDIRSQSSVDVSLQDIKNALGSLQGEGFLTVHGDIVKR 900

Qy        901 V 901
              |
Db        901 V 901
```

```
RESULT 3
ADY09589
ID   ADY09589 standard; protein; 881 AA.
XX
AC   ADY09589;
XX
DT   21-APR-2005  (first entry)
XX
DE   Plant full length insert polypeptide seqid 65404.
XX
KW   plant protectant; plant growth regulant; gene therapy; plant;
KW   recombinant DNA construct; physical array; plant breeding marker;
KW   cold tolerance; heat tolerance; drought tolerance; herbicide tolerance;
KW   extreme osmotic condition; pathogen tolerance; pest tolerance;
KW   growth rate; cell cycle pathway; disease resistance;
KW   galactomannan production; lignin production; plant growth regulator;
KW   yield; plant growth; plant development; seed oil; protein yield;
KW   protein content.
XX
OS   Unidentified.
XX
PN   US2004034888-A1.
XX
PD   19-FEB-2004.
XX
PF   28-APR-2003; 2003US-00425114.
XX
PR   06-MAY-1999;   99US-00304517.
PR   05-NOV-2001; 2001US-00985678.
XX
PA   (LIUJ/) LIU J.
PA   (ZHOU/) ZHOU Y.
PA   (KOVA/) KOVALIC D K.
PA   (SCRE/) SCREEN S E.
PA   (TABA/) TABASKA J E.
PA   (CAOY/) CAO Y.
XX
PI   Liu J,  Zhou Y,  Kovalic DK,  Screen SE,  Tabaska JE,  Cao Y;
XX
DR   WPI; 2004-180133/17.
XX
PT   New recombinant DNA construct, useful for improving plant tolerance to
PT   cold, heat, drought, herbicides, extreme osmotic conditions, pathogens or
PT   pests, for conferring increased resistance to plant disease, or for
PT   improving yield.
XX
PS   Claim 1; SEQ ID NO 65404; 15pp; English.
XX
CC   The invention describes a recombinant DNA construct comprising a
CC   polynucleotide consisting of a sequence encoding an amino acid sequence
CC   available in electronic form from the US patent office at
CC   ftp.seqdata.uspto.gov/sequence.html?DocID:2004034888. The polynucleotide
CC   of the invention are also useful in physical arrays of molecules and as
CC   plant breeding markers. The recombinant DNA construct is useful for
CC   improving plant tolerance to cold, heat, drought, herbicides, extreme
CC   osmotic conditions, pathogens or pests, for manipulating growth rate in
CC   plant cells by modification of the cell cycle pathway, for conferring
CC   increased resistance to plant disease, for producing galactomannan,
CC   lignin or plant growth regulators, for increasing the rate of homologous
CC   recombination in plants, for improving yield by modification of
CC   photosynthesis or carbohydrate, nitrogen or phosphorus use and/or uptake
CC   or by providing improved plant growth and development under at least one
CC   stress condition or for modifying seed oil or protein yield and/or
CC   content. This is the amino acid sequence of a plant full length insert
CC   polypeptide that can be used in the recombinant DNA construct of the
CC   invention.
XX
SQ   Sequence 881 AA;

  Query Match             55.3%;  Score 2550;  DB 8;  Length 881;
  Best Local Similarity   57.3%;  Pred. No. 9.5e-204;
  Matches  514;  Conservative  143;  Mismatches  176;  Indels   64;  Gaps   17;

Qy         20 SEGVSTPLPQVTSPSFDNAASPVAGRRAVRQTPTSAVRRRGRETDSARRRRSRSRSLGNS 79
              |:|  :  |  :|| :   :||:|       :|:   || |      ||||
Db         34 SDGGGSSPPPASSP-YGRPSSPLA---VTNSSPSQPTRRPG-----GRRRRGS------- 77

Qy         80 VYSSPYDAGTPGTPGTPVATPVYATPVGTPMGTPSFHRGTPQYKQRSELGSQGKPLHRRR 139
                :||| | :|   |       : ||       |   | ||     |  |: | |   |:
Db         78 --ASPY-ASSPSLGG-------FETP-------PHPGRRTP-----SGAGA-GAPRQPRQ 114

Qy        140 RSQSREPGHRSPSREPSADG-RPSESAEPDDTLGG-----EYAYVWGTNVNIPDVLRAIR 193
               |  | |   :||   | |   ||  |  |:| ||        :|||||::: ||  ||
Db        115 NSTGRFP--PTPSTPMSTDDVPPSSEAGDDETDGGGGVDATPVFVWGTNISVQDVNAAIL 172
```

```
Qy     194 RFLHNYRSSAHDL--------NSKYIQIIEETVERE-EDTLNIDMSDIYDHDPDLYAKIV 244
           ||| ::|      |              ||:: |    :| |   ::|::|   |::||||||||:|:|
Db     173 RFLRHFR-DPRDAGRVDPVMDEGKYMRAIHRILELEGGESLDVDAHDVFDHDPDLYSKMV 231

Qy     245 RYPLDIIPLLDTECQEVATSLLPTFEKHIEARPFNLKASVHMRELNPSDIDKLVSVKGMV 304
           ||||::: : |     ::    : | |||||: | :|||:|: :| ||||||:|:||:|||:
Db     232 RYPLEVLAIFDIVLMDLVARIEPLFEKHIQTRIYNLKSSICLRNLNPSDIEKMVSIKGMI 291

Qy     305 IRCSSIIPEIKGAFFKCLVCGHSPPLVTVVKGRVEEPTRCEKPECAARNAMSLIHNRCTF 364
           |||||:|||:| | |:||||||       | | :||| ||  |:| :| | |:|:|:||||| |
Db     292 IRCSSVIPELKEAVFRCLVCGFYSEPVMVDRGRVTEPHICQKEQCKATNSMTLVHNRCRF 351

Qy     365 ANKQIVRLQETPDAIPEGETPHTVSMCLYNTMVDAVKPGDRIEVTGVFKAMAVRVGPNQR 424
           ::|||:|||||| |||| ||||||: ::: :||| |||||:|:|||::|||:|:|| ||
Db     352 SDKQIIKLQETPDEIPEGGTPHTVSVLMHDKLVDAGKPGDRVEITGIYRAMSIRIGPTQR 411

Qy     425 TLRALYKTYIDCVHVKKSDRGRLQTEDPMEMDKENDMYAGYHESDTSEAANEAKIQKLKE 484
           |:::::||||||:|:|||:|: ||  || |::|  |               ||   |::||||
Db     412 TVKSIFKTYIDCLHIKKTDKSRLHVEDTMDIDNSN------ASKSTEEDFLSDKVEKLKE 465

Qy     485 LSKLPGIYDRLSRSLAPSIWELEDIKKGLLCQLFGGKAKKIPSGASFRGDINVLLVGDPG 544
           ||||| ||:||:|:|||||:||||:|:|:||||||||       |:||||||||||||:||||||
Db     466 LSKLPDIYERLTRSLAPNIWELDDVKRGLLCQLFGGNPLKLPSGASFRGDINILLVGDPG 525

Qy     545 TSKSQLLQYVHKIAPRGIYTSGRGSSAVGLTAYVTKDPETRETVLESGALVLSDRGICCI 604
           |||||||||:|||:|||||||||||||||||||||||||||| |||||||||||||||:|:|||
Db     526 TSKSQLLQYMHKLSPRGIYTSGRGSSAVGLTAYVTKDPETGETVLESGALVLSDKGVCCI 585

Qy     605 DEFDKMSDNARSMLHEVMEQQTVSVAKGGIIASLNARTSVLACANPSGSRYNARLSVIDN 664
           ||||||||||||||||||||||||:||:|| ||||||||||||||||||: |||| ||||||
Db     586 DEFDKMSDNARSMLHEVMEQQTVSIAKAGIIASLNARTSVLACANPTESRYNPRLSVIDN 645

Qy     665 IQLPPTLLSRFDLIYLMLDKPDEQNDRRLARHLVALHYENYEVSKQDALDLQTLTAYITY 724
           | | ||||||||||||:||| ||| |||||:|:|:||:||  : : : ||||| :||:|
Db     646 IHLAPTLLSRFDLIYLILDKADEQTDRRLAKHIVSLHFENPNLEELEVLDLQTLVSYISY 705

Qy     725 ARQHVHPTLSDEAAEDLINGYVEMRQKGNFPGSSKKVITATPRQLESMIRISEALARMRF 784
           ||::: | |||||||:| |||||::|| ||| |||||| ||:||:||:||||||||
Db     706 ARKYIQPQLSDEAAEELTRGYVEMRKRGNSPGSRKKVITATARQIESLIRLSEALARMRF 765

Qy     785 SEVVEKVDAAEAVRLLDVALQQSATDHATGTIDMDLITTGVSASERIRRANLLAALRELI 844
           |||||    | || |||:||:||||||||||||||||||| ||:||||||| || ||:|| | ||
Db     766 SEVVEVRDVVEAFRLLEVAMQQSATDHATGTIDMDLIMTGISASERQRRENLVAATRNLI 825

Qy     845 ADKISPGSSSGLKTSQLLEDIRSQSSVDVSLQDIKNALGSLQGEGFLTVHGDIVKRV 901
           |:|:  |   | ::   :|||::| |||::: : ::: |||:|   || : :||| |:||
Db     826 AEKMQLGGPS-MRMIELLEELRKQSSMEIHMHELRGALGTLMTEGAVVIHGDNVRRV 881
```

RESULT 4
ABJ26500

```
ID   ABJ26500 standard; protein; 1023 AA.
XX
AC   ABJ26500;
XX
DT   16-APR-2003  (first entry)
XX
DE   Aspergillus fumigatus essential gene protein #1158.
XX
KW   Fungicide; cytostatic; essential gene; Aspergillus fumigatus; infection;
KW   cancer; contamination; biofilm; antibody; immune response.
XX
OS   Aspergillus fumigatus.
XX
PN   WO200286090-A2.
XX
PD   31-OCT-2002.
XX
PF   23-APR-2002; 2002WO-US013142.
XX
PR   23-APR-2001; 2001US-0285697P.
PR   27-APR-2001; 2001US-0287066P.
PR   05-JUN-2001; 2001US-0295890P.
PR   09-JUL-2001; 2001US-0303899P.
PR   31-AUG-2001; 2001US-0316362P.
XX
PA   (ELIT-) ELITRA PHARM INC.
XX
PI   Jiang B,  Tishkoff D,  Zamudio C,  Eroshkin AM,  Hu W,  Lemieux SM;
XX
DR   WPI; 2003-093124/08.
XX
PT   New purified or isolated nucleic acids of essential genes of Aspergillus
PT   fumigatus, useful for treating or preventing infections by A. fumigatus,
PT   or for treating a non-infectious disease in a subject e.g. cancer.
XX
PS   Disclosure; Page; 175pp; English.
XX
CC   The invention relates to novel purified or isolated nucleic acids of
CC   essential genes of Aspergillus fumigatus. The isolated nucleic acids of
CC   the invention are used to treat or prevent infections by a pathogenic
CC   organism such as A. fumigatus, to treat a non-infectious disease in a
CC   subject (e.g. cancer), to prevent or contain contamination of an object
CC   by A. fumigatus, or to prevent or inhibit formation on a surface of a
CC   biofilm comprising A. fumigatus. The polynucleotides are useful for
CC   expressing recombinant protein for characterisation, screening or
CC   therapeutic use, as markers for host tissues in which the pathogenic
CC   organisms invade or reside, for comparing with the DNA sequence of A.
CC   fumigatus to identify duplicated genes or paralogues having the same or
CC   similar biochemical activity and/or function, for comparing with DNA
CC   sequences of other related or distant pathogenic organisms to identify
CC   potential orthologous essential or virulence genes, for selecting and
CC   making oligomers for attachment to a nucleic acid array for examination
CC   of expression patterns, for raising anti-protein antibodies, as an
CC   antigen to raise anti-DNA antibodies or to elicit another immune
CC   response, and for identifying polynucleotides encoding the other protein
CC   with which binding occurs or to identify inhibitors of the binding
CC   interaction. The polypeptides may be used to raise antibodies or to
CC   elicit immune response, as a reagent in assays designed to quantitatively
CC   determine levels of the protein in biological fluids, as a marker for
CC   host tissues in which pathogenic organism invade or reside, and to
CC   isolate correlative receptors or ligands in the case or virulence
CC   factors. This sequence represents a protein of one of the essential genes
CC   of Aspergillus fumigatus of the invention
XX
SQ   Sequence 1023 AA;

  Query Match             40.9%;  Score 1883;  DB 6;  Length 1023;
  Best Local Similarity   42.3%;  Pred. No. 9.6e-148;
  Matches  434;  Conservative  155;  Mismatches  286;  Indels  152;  Gaps   23;

Qy      12 VSSPYPSQSEGVSTPLPQVTSPSFDNAASPVAGRRAVRQTPTSAVRRRGRETDSARRRS 71
            :|||  |: |              |: | | | |     ||  ||:    |  :  |    |:
Db       1 MSSPASSRRRG---------RPAKDTATSSPARSTRSRQLQTSSPTPRAADEQSQATPRA 51

Qy      72 RSRSLGNSVY--SSPY-----------DAGTPGT----PGTPV----------ATPVYAT 104
             |  | :     |||                 | ||        | :|:            ||
Db      52 SRRLRGEAAVPSSSPMFFQSSPSKADSSAETPDVRMDEPSSPMRESSTMDEGDRTPRGNA 111

Qy     105 PV---GTP---MGTPSFHRGTPQYKQRSELGSQGKPLH------------RRRRSQSREP 146
           |     :|    |  : |  |   :  :||:: |      |                 |||
Db     112 PTMRDSSPIRYMSSSSPTRAQNRQSRRSDIPSSSSGLFVSSRPSIESNRAVSRRSDLHSG 171

Qy     147 GH-RSPSRE----------PSADGRPSESA-----EPD----DTLGGEYA-YVWGTNVNI 185
           |   ||:|             |: || |   |         ||      : |||       :||||::|
Db     172 GFLSSPNRRRRVFVDANGMPATDGDPRSDATFSNIHPDTSEAEALGGSSTRVIWGTNISI 231

Qy     186 PDVLRAIRRFLHN----YRSSAHDLNSKYIQIIEETVERE-------------DTLNID 228
            | : | : ||:|    || |        :|: :: |  |                 :||:|
Db     232 QDSMSAFKNFLYNFQTKYRLWAEGATEDETRIMGDSAEEREYISMLSTMRQLGVTSLNLD 291

Qy     229 MSDI--YDHDPDLYAKIVRYPLDIIPLLDTECQEVATSL--------------------- 265
             ::  |      |: ::  || :||||:|    ::|   |
Db     292 AKNLKAYPSTLKLWHQLHAYPQEIIPLMDQTVKDVMVELAIKEMERLRAQNQRNQNHNRG 351

Qy     266 ----------------------LPTFEKHIEAR-----PFNLKASVHMRELNPSDIDKL 297
```

```
Db        352 LSSGPAVPSSDALSETGRMPQNEIPDLVGEVETKAFKVLPFGLDSTVNMRDLDPADMDKL 411

Qy        298 VSVKGMVIRCSSIIPEIKGAFFKCLVCGHSPPLVTVVKGRVEEPTRCEKPECAARNAMSL 357

Db        412 VSIKGLVIRTTPIIPDMKEAFFRCQVCNHGVQ-VDIDRGKIAEPTECPRPVCKERNSMQL 470

Qy        358 IHNRCTFANKQIVRLQETPDAIPEGETPHTVSMCLYNTMVDAVKPGDRIEVTGVFKAMAV 417

Db        471 IHNRCVFADKQVIKLQETPDSIPDGQTPHSVSLCVYDELVDVCKAGDRVEVTGIFRCNPV 530

Qy        418 RVGPNQRTLRALYKTYIDCVHVKKSDRGRLQTEDPMEMDKENDMYAGYHESDTSEAANEA 477

Db        531 RVNPRQRTQKSLFKTYIDVLHVQKIDRKKLGIDVSTIEQELSEQAAGDAEQTRRLTAEEE 590

Qy        478 KIQKLKELSKLPGIYDRLSRSLAPSIWELEDIKKGLLCQLFGGKAKKIPSGAS--FRGDI 535

Db        591 --EKIKRTATRPDLYELLSRSLAPSIYEMDDVKKGILLQLFGGTNKTFQKGGNPRYRGDI 648

Qy        536 NVLLVGDPGTSKSQLLQYVHKIAPRGIYTSGRGSSAVGLTAYVTKDPETRETVLESGALV 595

Db        649 NILLCGDPSTSKSQLLRYVHKIAPRGVYTSGKGSSAVGLTAYVTRDPETRQMVLESGALV 708

Qy        596 LSDRGICCIDEFDKMSDNARSMLHEVMEQQTVSVAKGGIIASLNARTSVLACANPSGSRY 655

Db        709 LSDGGICCIDEFDKMNESTRSVLHEVMEQQTVSIAKAGIITTLNARTSILASANPIGSRY 768

Qy        656 NARLSVIDNIQLPPTLLSRFDLIYLMLDKPDEQNDRRLARHLVALHYEN--YEVSKQDAL 713

Db        769 NPNLPVPQNIDLPPTLLSRFDLVYLVLDRVDEQEDRRLAKHLVNMYLEDRPEHAAEQEIL 828

Qy        714 DLQTLTAYITYARQHVHPTLSDEAAEDLINGYVEMRQKGNFPGSSKKVITATPRQLESMI 773

Db        829 PIEFLTAYITYAKTKVHPVLTPAAGKALSDAYVNMRKLGDDIRSSDRRITATTRQLESMI 888

Qy        774 RISEALARMRFSEVVEKVDAAEAVRLLDVALQQSATDHATGTIDMDLITTGVSASERIRR 833

Db        889 RLSEAHARMRLSPEVTADDVEEAVRLIRSAIKQAATDSRTGLIDMSLLTEGTSASERRSR 948

Qy        834 ANLLAALRELIADKISPGSSSGLKTSQLLEDIRSQSSVDVSLQDIKNALGSLQGEGFLTV 893

Db        949 EALKRALLSVVDDLCSGGGAA--RWAEVFRILSENSSIEVDGAQFADAVRALEAEGAVSV 1006

Qy        894 HGDIVKR 900

Db       1007 VGEGARR 1013
```

RESULT 5
ABJ25900
ID ABJ25900 standard; protein; 874 AA.
XX
AC ABJ25900;
XX
DT 16-APR-2003 (first entry)
XX
DE Aspergillus fumigatus essential gene protein #558.
XX
KW Fungicide; cytostatic; essential gene; Aspergillus fumigatus; infection;
KW cancer; contamination; biofilm; antibody; immune response.
XX
OS Aspergillus fumigatus.
XX
PN WO200286090-A2.
XX
PD 31-OCT-2002.
XX
PF 23-APR-2002; 2002WO-US013142.
XX
PR 23-APR-2001; 2001US-0285697P.
PR 27-APR-2001; 2001US-0287066P.
PR 05-JUN-2001; 2001US-0295890P.
PR 09-JUL-2001; 2001US-0303899P.
PR 31-AUG-2001; 2001US-0316362P.
XX
PA (ELIT-) ELITRA PHARM INC.
XX
PI Jiang B, Tishkoff D, Zamudio C, Eroshkin AM, Hu W, Lemieux SM;
XX
DR WPI; 2003-093124/08.
XX
PT New purified or isolated nucleic acids of essential genes of Aspergillus
PT fumigatus, useful for treating or preventing infections by A. fumigatus,
PT or for treating a non-infectious disease in a subject e.g. cancer.
XX
PS Disclosure; Page; 175pp; English.
XX
CC The invention relates to novel purified or isolated nucleic acids of
CC essential genes of Aspergillus fumigatus. The isolated nucleic acids of
CC the invention are used to treat or prevent infections by a pathogenic
CC organism such as A. fumigatus, to treat a non-infectious disease in a
CC subject (e.g. cancer), to prevent or contain contamination of an object
CC by A. fumigatus, or to prevent or inhibit formation on a surface of a
CC biofilm comprising A. fumigatus. The polynucleotides are useful for
CC expressing recombinant protein for characterisation, screening or
CC therapeutic use, as markers for host tissues in which the pathogenic
CC organisms invade or reside, for comparing with the DNA sequence of A.
CC fumigatus to identify duplicated genes or paralogues having the same or
CC similar biochemical activity and/or function, for comparing with DNA
CC sequences of other related or distant pathogenic organisms to identify
CC potential orthologous essential or virulence genes, for selecting and
CC making oligomers for attachment to a nucleic acid array for examination
CC of expression patterns, for raising anti-protein antibodies, as an
CC antigen to raise anti-DNA antibodies or to elicit another immune
CC response, and for identifying polynucleotides encoding the other protein
CC with which binding occurs or to identify inhibitors of the binding
CC interaction. The polypeptides may be used to raise antibodies or to
CC elicit immune response, as a reagent in assays designed to quantitatively
CC determine levels of the protein in biological fluids, as a marker for
CC host tissues in which pathogenic organism invade or reside, and to
CC isolate correlative receptors or ligands in the case or virulence
CC factors. This sequence represents a protein of one of the essential genes
CC of Aspergillus fumigatus of the invention
XX
SQ Sequence 874 AA;

Query Match 40.4%; Score 1860; DB 6; Length 874;
Best Local Similarity 45.5%; Pred. No. 6.3e-146;
Matches 413; Conservative 146; Mismatches 244; Indels 104; Gaps 18;

```
Qy         51 TPTSAVRRRGRETDSARRRRSRSRSLGNSVYSSPYDAGTPGTPGTPVATPVYATPVGTPM 110

Db          5 SPTRAQNRQSRRSDIP------SSSSGLFVSSRP------------------------ 32

Qy        111 GTPSFHRGTPQYKQRSELGSQG--KPLHRRRRSQSREPGHRSPSREPSADGRPSESA--- 165

Db         33 -SIESNRAV---SRRSDLHSGGFLSSPNRRRRVFVDANG------MPATDGDPRSDATFS 82

Qy        166 --EPD----DTLGGEYA-YVWGTNVNIPDVLRAIRRFLHN----YRSSAHDLNSKYIQII 214

Db         83 NIHPDTSEAEALGGSSTRVIWGTNISIQDSMSAFKNFLYNFQTKYRLWAEGATEDETRIM 142

Qy        215 EETVEREE------------DTLNIDMSDI--YDHDPDLYAKIVRYPLDIIPLLDTECQ 259

Db        143 GDSAEEREYISMLSTMRQLGVTSLNLDAKNLKAYPSTLKLWHQLHAYPQEIIPLMDQTVK 202

Qy        260 EVATSL--------------------LPTFE-KHIEARPFNLKASVHMRELNPSDIDKL 297

Db        203 DVMVELAIKEMERLRAQNQRNQNHNRGLSSVETKAFKVLPFGLDSTVNMRDLDPADMDKL 262

Qy        298 VSVKGMVIRCSSIIPEIKGAFFKCLVCGHSPPLVTVVKGRVEEPTRCEKPECAARNAMSL 357

Db        263 VSIKGLVIRTTPIIPDMKEAFFRCQVCNHGVQ-VDIDRGKIAEPTECPRPVCKERNSMQL 321

Qy        358 IHNRCTFANKQIVRLQETPDAIPEGETPHTVSMCLYNTMVDAVKPGDRIEVTGVFKAMAV 417
```

```
Db        322 IHNRCVFADKQVIKLQETPDSIPDGQTPHSVSLCVYDELVDVCKAGDRVEVTGIFRCNPV 381

Qy        418 RVGPNQRTLRALYKTYIDCVHVKKSDRGRLQTEDPMEMDKENDMYAGYHESDTSEAANEA 477
              || | ||| ::|:||||| :||:| || :|  :        : ::  ||  |        | |
Db        382 RVNPRQRTQKSLFKTYIDVLHVQKIDRKKLGIDVSTIEQELSEQAAGDAEQTRRLTAEEE 441

Qy        478 KIQKLKELSKLPGIYDRLSRSLAPSIWELEDIKKGLLCQLFGGKAKKIPSGAS--FRGDI 535
                :|:|  :   | :|: ||||||||:|::|:|||:| |||||  |      | :  :||||
Db        442 --EKIKRTATRPDLYELLSRSLAPSIYEMDDVKKGILLQLFGGTNKTFQKGGNPRYRGDI 499

Qy        536 NVLLVGDPGTSKSQLLQYVHKIAPRGIYTSGRGSSAVGLTAYVTKDPETRETVLESGALV 595
              |:|| ||| |||||||:|||||||||:||||:||||||||||||:|||||: ||||||||
Db        500 NILLCGDPSTSKSQLLRYVHKIAPRGVYTSGKGSSAVGLTAYVTRDPETRQMVLESGALV 559

Qy        596 LSDRGICCIDEFDKMSDNARSMLHEVMEQQTVSVAKGGIIASLNARTSVLACANPSGSRY 655
              ||| |||||||||||:::  ||:|||||||||||:|| |||  :||||||:|| ||| ||||
Db        560 LSDGGICCIDEFDKMNESTRSVLHEVMEQQTVSIAKAGIITTLNARTSILASANPIGSRY 619

Qy        656 NARLSVIDNIQLPPTLLSRFDLIYLMLDKPDEQNDRRLARHLVALHYEN--YEVSKQDAL 713
               |  |  || |||||||||||:||:||: ||| ||||:|||  :: |:        ::|: |
Db        620 NPNLPVPQNIDLPPTLLSRFDLVYLVLDRVDEQEDRRLAKHLVNMYLEDRPEHAAEQEIL 679

Qy        714 DLQTLTAYITYARQHVHPTLSDEAAEDLINGYVEMRQKGNFPGSSKKVITATPRQLESMI 773
               :: |||||||:  ||| |:  | : | : || ||: |:   || : |||| ||||||||
Db        680 PIEFLTAYITYAKTKVHPVLTPAAGKALSDAYVNMRKLGDDIRSSDRRITATTRQLESMI 739

Qy        774 RISEALARMRFSEVVEKVDAAEAVRLLDVALQQSATDHATGTIDMDLITTGVSASERIRR 833
              |:||| |||| |  |   | |||||:  |::|:|||  || ||| |:| | ||||| |
Db        740 RLSEAHARMRLSPEVTADDVEEAVRLIRSAIKQAATDSRTGLIDMSLLTEGTSASERRSR 799

Qy        834 ANLLAALRELIADKISPGSSSGLKTSQLLEDIRSQSSVDVSLQDIKNALGSLQGEGFLTV 893
               | ||  :: | | | :: : ::: :   ||::|        |: :|: || ::|
Db        800 EALKRALLSVVDDLCSGGGAA--RWAEVFRILSENSSIEVDGAQFADAVRALEAEGAVSV 857

Qy        894 HGDIVKR 900
              |:  :|
Db        858 VGEGARR 864
```

RESULT 6
ADB70288

```
ID   ADB70288 standard; protein; 989 AA.
XX
AC   ADB70288;
XX
DT   04-DEC-2003  (first entry)
XX
DE   C. neoformans amino acid sequence SEQ ID NO:3332.
XX
KW   fungicide; gene therapy; infection.
XX
OS   Cryptococcus neoformans.
XX
PN   WO2003052076-A2.
XX
PD   26-JUN-2003.
XX
PF   17-DEC-2002; 2002WO-US040225.
XX
PR   17-DEC-2001; 2001US-0341261P.
XX
PA   (ELIT-) ELITRA PHARM INC.
XX
PI   Zamudio C,  Eroshkin AM;
XX
DR   WPI; 2003-533017/50.
DR   N-PSDB; ADB69205.
XX
PT   New nucleic acid, useful for preparing a composition for treating an
PT   infection caused by Cryptococcus neoformans.
XX
PS   Claim 9; SEQ ID NO 3332; 136pp; English.
XX
CC   The invention relates to a novel purified or isolated Cryptococcus
CC   neoformans nucleic acid molecule comprising a sequence encoding a
CC   polypeptide comprising a sequence not given in the specification. A
CC   polynucleotide of the invention has fungicide activity, and may have a
CC   use in gene therapy. The nucleic acid is useful for preparing a
CC   composition for treating an infection caused by Cryptococcus neoformans.
CC   The present sequence represents a C. neoformans sequence of the
CC   invention. Note: The sequence data for this patent is not represented in
CC   the printed specification, but was obtained in electronic format directly
CC   from WIPO at ftp.wipo.int/pub/published_pat_sequences.
XX
SQ   Sequence 989 AA;

  Query Match             39.8%;  Score 1835.5;  DB 7;  Length 989;
  Best Local Similarity   43.8%;  Pred. No. 8.7e-144;
  Matches  409;  Conservative  147;  Mismatches  246;  Indels  131;  Gaps   20;

Qy         47 AVRQTPTSAVRR-----RGRETDSARRRRSRSRSLGNSVYSSP--------------YD 86
              |: |:  ||::           ||       |  ||  |        :  :|||
Db          4 ALLQSDESAMKHHCSSLRGSSGGSTPRRARRGDIHSSFPHSSPSLARRQAQVPADSLSLR 63

Qy         87 AGTPGTPGTPVATPVYATPVGTPMGTPSFHRGTPQYKQRSELGSQGKPLHRRRRSQSREP 146
              | :||                 ||| ||  |||:         ||    |      :||
Db         64 ASSPG-----------------MGTDSF--GTPRV-----FGSAAPTLSAVAQSQL--- 95

Qy        147 GHRSPSREPSADGRPSESAEPDDTLGGEYAYVWGTNVNIPDVLRAIRRFL---------- 196
                       |||        | : |     ::||| ::: : :    | ||
Db         96 ---------GADG---------DDIDGMVKFIWGTTISLQESMNLFRDFLRGFKPKYRAV 137

Qy        197 HNYRSSAHDLNS---------KYIQIIEETVE-------------REEDTLNIDMSDIYD 234
              :|   | |  |             |  :  | |                    |  ||:|  ::
Db        138 YNAEQSRHAEESGGVAPPPMTLYDNLSAERAEVPLYETYLNRLRLTGETNLNLDALNLLA 197

Qy        235 HDP--DLYAKIVRYPLDIIPLLDTECQEVAT-----------------SLLPTFEKHI-E 274
              : |    || ::| || ::||::|    ::|                       :|   |  : :
Db        198 YRPTKKLYQQLVNYPQEVIPIMDQVLRDVMIELGHEELEKAQTKFAEGNLSQDVESRVYK 257

Qy        275 ARPFNLKASVHMRELNPSDIDKLVSVKGMVIRCSSIIPEIKGAFFKCLVCGHSPPLVTVV 334
               |||  : :|:||:||| ||||:|||:||| : :||::  |||:|||| |:        :
Db        258 VRPFGGEKTVNMRDLNPGDTDKLVTVKGLVIRATPVIPDMTTAFFRCLVCQHTVQ-ADID 316

Qy        335 KGRVEEPTRCEKPECAARNAMSLIHNRCTFANKQIVRLQETPDAIPEGETPHTVSMCLYN 394
              :||: || || :  | :    ||||||  :||::|||||||||:|:|:||||||:|:|:
Db        317 RGRISEPERCPRDVCGSTGTMSLIHNRSEFTDKQVIRLQETPDAVPDGQTPHTVSLCVYD 376

Qy        395 TMVDAVKPGDRIEVTGVFKAMAVRVGPNQRTLRALYKTYIDCVHVKKSDRGRLQTEDPME 454
               :|| ||||||: :||:|::: ||| | ||::::||||:| ||||::: |:  ||
Db        377 ELVDLVKPGDRVIITGIFRSIPVRVNPRQRSIKSLYKTYLDVVHVKRTNTARMGF-DPST 435

Qy        455 MDKENDMYAGYHESDTSEAANEAKIQKLKELSKLPGIYDRLSRSLAPSIWELEDIKKGLL 514
                |:   |  |     || | : |:: |||  | :|: |: |||||:||||:|||:|
Db        436 RAGESKP-PGDLGSPVRSAAAEME-QRIIELSNHPDLYNILASSLAPSIYELEDVKKGIL 493

Qy        515 CQLFGGKAKKIPSGAS-----FRGDINVLLVGDPGTSKSQLLQYVHKIAPRGIYTSGRGS 569
               |||| | |  |          :||||||:|||||||||||:||||||||||:||||:||
Db        494 LQLFGGTNKSIARGGGGGGPRYRGDINVLMVGDPGTSKSQILQYVHKIAPRGVYTSGKGS 553

Qy        570 SAVGLTAYVTKDPETRETVLESGALVLSDRGICCIDEFDKMSDNARSMLHEVMEQQTVSV 629
              |||||||||:||:::: |||||||||| |:|||||||||||    ||:||||||||||:
Db        554 SAVGLTAYVTRDPDSKQLVLESGALVLSDGGVCCIDEFDKMSDATRSVLHEVMEQQTVSI 613

Qy        630 AKGGIIASLNARTSVLACANPSGSRYNARLSVIDNIQLPPTLLSRFDLIYLMLDKPDEQN 689
              || ||| :||||||:|| ||| |||:  | :  || |||||:||||||:||:||| || |
Db        614 AKAGIITTLNARTSILAAANPINSRYDPNLPIPANIDLPPTLISRFDLLYLVLDKVDEVN 673

Qy        690 DRRLARHLVALHYENYEVSKQD-ALDLQTLTAYITYARQHVHPTLSDEAAEDLINGYVEM 748
              ||:||:||| |:  : |     | : |||||:||||||  :|| |:: |:| |:  ||||
Db        674 DRKLAKHLVGLYLSDVEDQPADNIIPLQTLTSYITYARSKIHPVLTEGASEALVQAYVEM 733

Qy        749 RQKGNFPGSSKKVITATPRQLESMIRISEALARMRFSEVVEKVDAAEAVRLLDVALQQSA 808
              |: |      : :| ||||||||||: || |||| |:  ||: |  |||||:  ||::||
Db        734 RKAGMDSRTQEKRITATTRQLESMIRLGEAHARMRLSDRVEEEDIREAVRLIKSALRESA 793
```

```
Qy        809 TDHATGTIDMDLITTGVSASERIRRANLLAALRELIADKISPGSSSGLKTSQLLEDIRSQ 868
              ||  || ||:||| ||     :  |   ||:|    : :|: :|        | |:: : ::::: |
Db        794 TDPLTGQIDLDLINTGAGQTMRRARADLKREVIKLVVEK---ARSQGIRWAAVIDELNKQ 850

Qy        869 SSVDVSLQDIKNALGSLQGEGFLTVHGDIVKRV 901
              |||  |          :    |:  |   :  |  |:   :|:
Db        851 SSVPVDHAQFAEIVRELEEESIVKVMGERERRI 883


RESULT 7
ADN21148
ID   ADN21148 standard; protein; 960 AA.
XX
AC   ADN21148;
XX
DT   02-DEC-2004  (first entry)
XX
DE   Bacterial polypeptide #3801.
XX
KW   Recombinant DNA construct; transformed plant; improved plant property;
KW   cold tolerance; heat tolerance; drought tolerance; herbicide; osmosis;
KW   pathogen tolerance; pest tolerance; plant disease resistance;
KW   cell cycle pathway modification; plant growth regulator;
KW   homologous recombination; seed oil yield; protein yield; carbohydrate;
KW   nitrogen; phosphorus; photosynthesis; lignin; galactomannan;
KW   bacterial polypeptide.
XX
OS   Bacteria.
XX
PN   US2003233675-A1.
XX
PD   18-DEC-2003.
XX
PF   20-FEB-2003; 2003US-00369493.
XX
PR   21-FEB-2002; 2002US-0360039P.
XX
PA   (CAOY/) CAO Y.
PA   (HINK/) HINKLE G J.
PA   (SLAT/) SLATER S C.
PA   (CHEN/) CHEN X.
PA   (GOLD/) GOLDMAN B S.
XX
PI   Cao Y,  Hinkle GJ,  Slater SC,  Chen X,  Goldman BS;
XX
DR   WPI; 2004-061375/06.
XX
PT   New recombinant DNA construct comprising a promoter positioned to provide
PT   for expression of a polynucleotide encoding a polypeptide from a
PT   microbial source, useful for producing plants with improved properties.
XX
PS   Claim 1; SEQ ID NO 3801; 122pp; English.
XX
CC   The invention relates to a recombinant DNA construct comprising a
CC   promoter functional in a plant cell, where the promoter is positioned to
CC   provide for expression of a polynucleotide encoding a polypeptide from a
CC   microbial source. The invention also relates to a transformed plant
CC   comprising the recombinant DNA construct and a method of producing a
CC   transformed plant having an improved property. The plant is a crop plant
CC   such as maize or soybean. The method of producing a transformed plant
CC   having an improved property comprises transforming a plant with the
CC   recombinant DNA construct and growing the transformed plant, where the
CC   polynucleotide or polypeptide is useful for improving plant properties.
CC   The recombinant DNA construct is useful for producing plants with
CC   improved plant properties, e.g. improved cold, heat or drought tolerance,
CC   tolerance to herbicides, extreme osmotic conditions, pathogens or pests,
CC   increased resistance to plant disease, better growth rate by modification
CC   of the cell cycle pathway with plant growth regulators, increased rate of
CC   homologous recombination, modified seed oil or protein yield and/or
CC   content, improved yield by modification of carbohydrate, nitrogen or
CC   phosphorus use and/or uptake, by modification of photosynthesis or by
CC   providing improved plant growth and development under at least one stress
CC   condition, improved lignin production or improved galactomannan
CC   production. This sequence represents a bacterial polypeptide used in the
CC   scope of the invention. Note: The sequence data for this patent did not
CC   form part of the printed specification but was obtained in electronic
CC   format from USPTO at seqdata.uspto.gov/sequence.html.
XX
SQ   Sequence 960 AA;

  Query Match             39.8%;  Score 1835;  DB 8;  Length 960;
  Best Local Similarity   41.2%;  Pred. No. 9.1e-144;
  Matches  412;  Conservative  163;  Mismatches  248;  Indels  178;  Gaps   23;

Qy         24 STPLPQVTSPSFD-------NAASPVAGRRAVRQTPTSAVRRRGRETDSARRRRSRSRSL 76
              |:|| | |     |        |  | :  | ||                :::: :
Db          1 SSPLRQQTETQSDADRTPRANGRSQLIGGMAVNQ---------------KKKKKDYTECX 45

Qy         77 GNSVYSSPYDAGTPGTPGTPVATPVYATPVGTPMGTPSFHRGTPQYKQRSE-----LGSQ 131
               :| |||    :  :|| :                     | |  |||     : ||
Db         46 PHSADSSPIRYASSSSPGRQL---------------------TQQSDLRSESSQLFVSSQ 84

Qy        132 GKPLHRRRRSQSREPGHRSPSR----------------EPSADGRPSESAEPD----DTL 171
                | ||       |:|::                   | :|      :  |:     |  |
Db         85 RSVAGRSRRGDINGDPLRTPAQIPRRIILDDAGRVIRDAPGSDANSFVTNNPNTSEADAL 144

Qy        172 GGE-YAYVWGTNVNIPDVLRAIRRFLHN----YRSSAH-----------DLNSK-YIQII 214
              ||:    |||| :::   |    :  ||  |        || |       :|| |  : :
Db        145 GGQSQGLVWGTTISLDDSFSAFKDFLRNFTRKYRMWADGADEAETIGHPDADSKPYWEAL 204

Qy        215 EETVEREEDTLNIDMSDI--YDHDPDLYAKIVRYPLDIIPLLDT---EC------QEVAT 263
              |  :    :  | :|: |:  |          |: :   || :|||::|     :|      :|:|:
Db        205 ENMLLLGTNKLYLDLRDLKSYPRTLKLWHQAQHYPTEIIPVMDQCVHDCMMELAQKEMAS 264

Qy        264 -------------------------------------SLLPTFEKHIE-----ARPFNLK 281
                                                   :  || |  :          ||: |
Db        265 QRASQNSRTAPGASQSSEPNFPSSERSEEPPTPRPAQTAAPTIEDQVSQMAYVVRPWGLD 324

Qy        282 ASVHMRELNPSDIDKLVSVKGMVIRCSSIIPEIKGAFFKCLVCGHSPPLVTVV--KGRVE 339
                ::|:|||||:|||||:||:|||  : :||::|  |||||  |||||    :||    :|::
Db        325 KITNLRDLNPSDMDKLVSIKGLVIRTTPVIPDMKDAFFKCSVCGHS---ITVQLDRGKIR 381

Qy        340 EPTRCEKPECAARNAMSLIHNRCTFANKQIVRLQETPDAIPEGETPHTVSMCLYNTMVDA 399
              ||| |  :   ||::|:|  :||||   |:|||:::|||||  :|  |:|||:|:|:|| :||
Db        382 EPTECPRARCASKNSMQIIHNRCAFEDKQVIKLQETPDNVPAGQTPHSVSVCVYNELVDF 441

Qy        400 VKPGDRIEVTGVFKAMAVRVGPNQRTLRALYKTYIDCVHVKKSDRGRLQTEDPMEMD--K 457
               | |||:|:||:||   |||  |  ||::::::|||:|  |||:|  ||  |: :  ||  :|  :
Db        442 CKAGDRVELTGIFKVTPVRVNPRMRTVKSVHKTYVDVVHVQKVDRKRMGS-DPSTLDLAE 500

Qy        458 ENDMYAGYHESDTSEAANEAKIQKLKELSKLPGIYDRLSRSLAPSIWELEDIKKGLLCQL 517
              | : :|        |        :    : :::||  :    | |||  |||||||||:|::|:|||:|  ||
Db        501 EEEAHANGQSMDEVRKVSPDEEERIKETAARPDIYDLLSRSLAPSIYEMDDVKKGILLQL 560

Qy        518 FGGKAKKIPSGAS--FRGDINVLLVGDPGTSKSQLLQYVHKIAPRGIYTSGRGSSAVGLT 575
              |||  |      | |   :|||||||| |||  |||||||| |||:||||||:|||||:||||||||
Db        561 FGGTNKTFEKGGSPKYRGDINVLLCGDPSTSKSQLLSYVHRIAPRGVYTSGKGSSAVGLT 620

Qy        576 AYVTKDPETRETVLESGALVLSDRGICCIDEFDKMSDNARSMLHEVMEQQTVSVAKGGII 635
              ||||:|||:|:  |||||||||| |:|||||||||:::  ||:|||||||||||||||||| |||
Db        621 AYVTRDPESRQLVLESGALVLSDGGVCCIDEFDKMNESTRSVLHEVMEQQTVSVAKAGII 680

Qy        636 ASLNARTSVLACANPSGSRYNARLSVIDNIQLPPTLLSRFDLIYLMLDKPDEQNDRRLAR 695
               :||||||:|| ||| |||||   |||  ||  |||||||||||:||:||:  ||:||:||||
Db        681 TTLNARTSILASANPIGSRYNPDLSVPQNIDLPPTLLSRFDLVYLILDRVDEKNDQRLAR 740

Qy        696 HLVALHYENYEVSKQDALDL----------------------------QTLTAYITYARQH 728
              ||::::  |:     | | | |:                           :  ||:||:||| |
Db        741 HLLSMYLEDKPESAQQANDVLVSFSPSVNPHEDEPQTNQLPLKKTARRVLTSYISYARSH 800

Qy        729 VHPTLSDEAAEDLINGYVEMRQKGNFPGSSKKVITATPRQLESMIRISEALARMRFSEVV 788
               :|| |: ||   :|::  |||||: |         :::| |||| ||||||||::|| |:|| |: |
Db        801 IHPALTPEAGRELVDAYVEMRKLGQDVRAAEKRITATTRQLESMIRLAEAHAKMRLSQTV 860

Qy        789 EKVDAAEAVRLLDVALQQSATDHATGTIDMDLITTGVSASERIRRANLLAALRELIADKI 848
               : |   |||||:   ||: :|||  :  | |||  |:|  | ||:||  |:|::  |:  |: :
```

```
Db        861 TRDDVREAVRLIKSALKTAATD-SQGRIDMSLLTEGTSAAERQRKADMKDAVIRLLDEMT 919

Qy        849 SPGSSSGLKTSQLLEDIRSQSSVDVSLQDIKNALGSLQGEG 889

Db        920 SGGQV--VRYSEVARRLGEGAGVQVEPAEFAEVMRALEMEG 958
```

RESULT 8
ADS44081
ID ADS44081 standard; protein; 933 AA.
XX
AC ADS44081;
XX
DT 02-DEC-2004 (first entry)
XX
DE Bacterial polypeptide #22511.
XX
KW Recombinant DNA construct; transformed plant; improved plant property;
KW cold tolerance; heat tolerance; drought tolerance; herbicide; osmosis;
KW pathogen tolerance; pest tolerance; plant disease resistance;
KW cell cycle pathway modification; plant growth regulator;
KW homologous recombination; seed oil yield; protein yield; carbohydrate;
KW nitrogen; phosphorus; photosynthesis; lignin; galactomannan;
KW bacterial polypeptide.
XX
OS Bacteria.
XX
PN US2003233675-A1.
XX
PD 18-DEC-2003.
XX
PF 20-FEB-2003; 2003US-00369493.
XX
PR 21-FEB-2002; 2002US-0360039P.
XX
PA (CAOY/) CAO Y.
PA (HINK/) HINKLE G J.
PA (SLAT/) SLATER S C.
PA (CHEN/) CHEN X.
PA (GOLD/) GOLDMAN B S.
XX
PI Cao Y, Hinkle GJ, Slater SC, Chen X, Goldman BS;
XX
DR WPI; 2004-061375/06.
XX
PT New recombinant DNA construct comprising a promoter positioned to provide
PT for expression of a polynucleotide encoding a polypeptide from a
PT microbial source, useful for producing plants with improved properties.
XX
PS Claim 1; SEQ ID NO 22511; 122pp; English.
XX
CC The invention relates to a recombinant DNA construct comprising a
CC promoter functional in a plant cell, where the promoter is positioned to
CC provide for expression of a polynucleotide encoding a polypeptide from a
CC microbial source. The invention also relates to a transformed plant
CC comprising the recombinant DNA construct and a method of producing a
CC transformed plant having an improved property. The plant is a crop plant
CC such as maize or soybean. The method of producing a transformed plant
CC having an improved property comprises transforming a plant with the
CC recombinant DNA construct and growing the transformed plant, where the
CC polynucleotide or polypeptide is useful for improving plant properties.
CC The recombinant DNA construct is useful for producing plants with
CC improved plant properties, e.g. improved cold, heat or drought tolerance,
CC tolerance to herbicides, extreme osmotic conditions, pathogens or pests,
CC increased resistance to plant disease, better growth rate by modification
CC of the cell cycle pathway with plant growth regulators, increased rate of
CC homologous recombination, modified seed oil or protein yield and/or
CC content, improved yield by modification of carbohydrate, nitrogen or
CC phosphorus use and/or uptake, by modification of photosynthesis or by
CC providing improved plant growth and development under at least one stress
CC condition, improved lignin production or improved galactomannan
CC production. This sequence represents a bacterial polypeptide used in the
CC scope of the invention. Note: The sequence data for this patent did not
CC form part of the printed specification but was obtained in electronic
CC format from USPTO at seqdata.uspto.gov/sequence.html.
XX
SQ Sequence 933 AA;

Query Match 39.8%; Score 1833; DB 8; Length 933;
Best Local Similarity 40.9%; Pred. No. 1.3e-143;
Matches 393; Conservative 178; Mismatches 281; Indels 110; Gaps 14;

```
Qy          2 ENNDALDIGAVSSPYPSQSEGVSTPLPQVTSPSFDNAASPVAG----------------- 44

Db         11 EDNNS------SSPVVPNPDSVP---PQLSSPALFYSSSSSQGDIYGRNNSQNLSQGEGN 61

Qy         45 -RRAVRQTP---TSAVRRRGRETDSARRRRSRSRSLGNSVYSSPYDAGTPGTPGTPVATP 100

Db         62 IRAAIGSSPLNFPSSSQRQNSDVFQSQGRQGRIRSSASASGRSRYHSDLRSDRALPTSS- 120

Qy        101 VYATPVGTPMGTPSFHRGTPQYKQRSELGSQGKPLHRRRRSQSREPGHRSPSREPSADGR 160

Db        121 -------------------------SSLGRNGQNRVHMRRNDIHTSDLSSPRRIVDFDTR 155

Qy        161 --------------PSESAEPDDTLGGEYAYVWGTNVNIPDVLRAIRRFLHNYR------ 200

Db        156 SGVNTLDTSSSSAPPSEASEP-------LRIIWGTNVSIQECTTNFRNFLMSFKYKFRKI 208

Qy        201 --------SSAHDLNSKYIQIIEETVEREEDTLNIDMSDI--YDHDPDLYAKIVRYPLDI 250

Db        209 LDEREEFINNTTDEELYYIKQLNEMRELGTSNLNLDARNLLAYKQTEDLYHQLLNYPQEV 268

Qy        251 IPLLDTECQEVATSL---------LPTFE-KHIEARPFNLKASVHMRELNPSDIDKLVSV 300

Db        269 ISIMDQTIKDCMVSLIVDNNLDYDLDEIETKFYKVRPYNVGSCKGMRELNPNDIDKLINL 328

Qy        301 KGMVIRCSSIIPEIKGAFFKCLVCGHSPPLVTVVKGRVEEPTRCEKPECAARNAMSLIHN 360

Db        329 KGLVLRSTPVIPDMKVAFFKCNVCDHT-MAVEIDRGVIQEPARCERIDCNEPNSMSLIHN 387

Qy        361 RCTFANKQIVRLQETPDAIPEGETPHTVSMCLYNTMVDAVKPGDRIEVTGVFKAMAVRVG 420

Db        388 RCSFADKQVIKLQETPDFVPDGQTPHSISLCVYDELVDSCRAGDRIEVTGTFRSIPIRAN 447

Qy        421 PNQRTLRALYKTYIDCVHVKKSDRGRLQTEDPMEMDKENDMYAGYHESDTSEAANEAKIQ 480

Db        448 SRQRVLKSLYKTYVDVVHVKKVSDKRLDVDTSTIEQELMQNKVDHNEVEEVRQITDQDLA 507

Qy        481 KLKELSKLPGIYDRLSRSLAPSIWELEDIKKGLLCQLFGGKAKKIPSGASFRGDINVLLV 540

Db        508 KIREVAAREDLYSLLARSIAPSIYELEDVKKGILLQLFGGTNKTFTKGGRYRGDINILLC 567

Qy        541 GDPGTSKSQLLQYVHKIAPRGIYTSGRGSSAVGLTAYVTKDPETRETVLESGALVLSDRG 600

Db        568 GDPSTSKSQILQYVHKITPRGVYTSGKGSSAVGLTAYITRDVDTKQLVLESGALVLSDGG 627

Qy        601 ICCIDEFDKMSDNARSMLHEVMEQQTVSVAKGGIIASLNARTSVLACANPSGSRYNARLS 660

Db        628 VCCIDEFDKMSDSTRSVLHEVMEQQTISIAKAGIITTLNARSSILASANPIGSRYNPNLP 687

Qy        661 VIDNIQLPPTLLSRFDLIYLMLDKPDEQNDRRLARHLVALHYEN--YEVSKQDALDLQTL 718

Db        688 VTENIDLPPPLLSRFDLVYLVLDKVDEKNDRELAKHLTNLYLEDKPEHISQDDVLPVEFL 747

Qy        719 TAYITYARQHVHPTLSDEAAEDLINGYVEMRQKGNFPGSSKKVITATPRQLESMIRISEA 778

Db        748 TMYISYAKEHIHPIITEAAKTELVRAYVGMRKMGDDSRSDEKRITATTRQLESMIRLAEA 807

Qy        779 LARMRFSEVVEKVDAAEAVRLLDVALQQSATDHATGTIDMDLITTGVSASERIRRANLLA 838

Db        808 HAKMKLKNVVELEDVQEAVRLIRSAIKDYATDPKTGKIDMNLVQTGKSVIQRKLQEDLSR 867

Qy        839 ALRELIADKISPGSSSGLKTSQLLEDIRSQSSVDVSLQDIKNALGSLQGEGFLTVHGDIV 898

Db        868 EIMNVLKDQ----ASDSMSFNELIKQINEHSQDRVESSDIQEALSRLQQEDKVIVLGEGV 923
```

```
Qy        899 KR 900
              :|
Db        924 RR 925
```

RESULT 9
ADD18710
ID   ADD18710 standard; protein; 923 AA.
XX
AC   ADD18710;
XX
DT   15-JAN-2004 (first entry)
XX
DE   Human disease related protein SeqID141.
XX
KW   human; disease state; cytostatic; antiinflammatory; ophthalmological;
KW   antiarteriosclerotic; vulnerary; gene therapy;
KW   hypoxia-regulated condition; tumourigenesis; angiogenesis; apoptosis;
KW   inflammation; erythropoiesis; glycolysis; gluconeogenesis;
KW   glucose transportation; catecholamine synthesis; iron transport;
KW   nitric oxide synthesis; cancer; ischaemic condition; reperfusion injury;
KW   retinopathy; neonatal stress; pre-eclampsia; atherosclerosis;
KW   inflammatory condition; wound healing.
XX
OS   Homo sapiens.
XX
PN   WO2003018621-A2.
XX
PD   06-MAR-2003.
XX
PF   23-AUG-2002; 2002WO-GB003892.
XX
PR   23-AUG-2001; 2001GB-00020558.
PR   05-OCT-2001; 2001GB-00024037.
XX
PA   (OXFO-) OXFORD BIOMEDICA UK LTD.
XX
PI   Kingsman SM, White J, Ward NR, Harris RA, Naylor S, Mundy CR;
XX
DR   WPI; 2003-290046/28.
DR   N-PSDB; ADD18711.
XX
PT   New substantially purified polypeptide, useful for diagnosing or treating
PT   a hypoxia-regulated condition, such as cancer, ischemia, reperfusion
PT   injury, retinopathy, pre-eclampsia, atherosclerosis, inflammation, or
PT   wound healing.
XX
PS   Claim 25; SEQ ID NO 141; 424pp; English.
XX
CC   This invention relates to novel human genes and gene product which are
CC   implicated in certain disease states. Compounds which modulate the
CC   proteins of the invention may have cytostatic, antiinflammatory,
CC   ophthalmological, antiarteriosclerotic or vulnerary activities. The
CC   sequences of the invention may be useful for gene therapy. The invention
CC   may be useful for diagnosing or treating a hypoxia-regulated condition,
CC   such as tumourigenesis, angiogenesis, apoptosis, inflammation,
CC   erythropoiesis, or the biological response to hypoxia conditions
CC   including processes such as glycolysis, gluconeogenesis, glucose
CC   transportation, catecholamine synthesis, iron transport or nitric oxide
CC   synthesis. The disease includes cancer, ischaemic conditions, reperfusion
CC   injury, retinopathy, neonatal stress, pre-eclampsia, atherosclerosis,
CC   inflammatory conditions or wound healing. The present sequence is that of
CC   a disease related protein of the invention.
XX
SQ   Sequence 923 AA;

```
  Query Match             39.4%;  Score 1817.5;  DB 7;  Length 923;
  Best Local Similarity   43.7%;  Pred. No. 2.5e-142;
  Matches  404;  Conservative  159;  Mismatches  275;  Indels   86;  Gaps   19;

Qy         18 SQSEGVSTPLP-QVTSPSFDNAASPVA-----GRRAVRQTPTSAVRRR-GRETDSARRRR 70
              |:  | :   | :  :||     :|| :        | |  | ||     |    | : | |||
Db         42 SREAGPARACPCRAGTPS--TMSSPASTPSRRGSRRGRATPAQTPRSEDARSSPSQRRRG 99

Qy         71 SRSRSLG-------------------NSVYSSPYDAGTPGTP-GTPVATPV-YATPVGTP 109
              | | |                       : ::|||     :     |     |::|: | ||
Db        100 EDSTSTGELQPMPTSPGVDLQSTAAQDVLFSSPPQMHSSAIPLDFDVSSPLTYGTPSSRV 159

Qy        110 MGTP-SFHRGTPQYKQRSELGSQGKPLHRRRRSQSREPGHRSPSREPSADGRPSESAEPD 168
              ||| |  |||| :|| :|||  | |      :|                   ||  :|
Db        160 EGTPRSGVRGTP-VRQRPDLGSAQKGLQVDLQS---------------DGAAAEDIVAS 202

Qy        169 DTLGGEYAYVWGTNVNIPDVLRAIRRFLHNY--------RSSAHDLNSK-YIQIIEETVE 219
              :   |:   :|||:||:                :|||    :          |:    |:| :  |
Db        203 EQSLGQKLVIWGTDVNVAACKENFQRFLQRFIDPLAKEEENVGIDITEPLYMQRLGEINV 262

Qy        220 REEDTLNIDMSDIYDHDPDLYAKIVRYPLDIIPLLDTECQEVATSLLP--TFEKHIEARP 277
                 ||::    |     | :||  ::: ||  :||    |      |:        |    |   |: ||
Db        263 IGEQFLNVNCEHIKSFDKNLYRQLISYPQEVIPTFDMAVNEIFFDRYPDSILEHQIQVRP 322

Qy        278 FNLKASVHMRELNPSDIDKLVSVKGMVIRCSSIIPEIKGAFFKCLVCGHSPPLVTVVKGR 337
              ||    : :||  |||  |||:|:::  |||||  |  :|||::  |||:|  || |:     |  : :||
Db        323 FNALKTKNMRNLNPEDIDQLITISGMVIRTSQLIPEMQEAFFQCQVCAHT-TRVEMDRGR 381

Qy        338 VEEPTRCEKPECAARNAMSLIHNRCTFANKQIVRLQETPDAIPEGETPHTVSMCLYNTMV 397
              : ||:  | :    |      ::|:||||||   |::||:::||||:|: :|  | |:|||||  :    :| :|
Db        382 IAEPSVCGR--CHTTHSMALIHNRSLFSDKQMIKLQESPEDMPAGQTPHTVILFAHNDLV 439

Qy        398 DAVKPGDRIEVTGVFKAMAVRVGPNQRTLRALYKTYIDCVHVKKSDRGRLQTEDPMEMDK 457
              | |:||||: |||:::|: :|| |        ::::|||:||  :| :|:|   ||
Db        440 DKVQPGDRVNVTGIYRAVPIRVNPRVSNVKSVYKTHIDVIHYRKTDAKRLH--------- 490

Qy        458 ENDMYAGYHESDTSEAANEAKIQKLKELSKLPGIYDRLSRSLAPSIWELEDIKKGLLCQL 517
                    |   |        :    :|  :::  ||||:  |  ||:||:   :|||||:|  ||||||:|  ||
Db        491 ------GLDEEAEQKLFSEKRVELLKELSRKPDIYERLASALAPSIYEHEDIKKGILLQL 544

Qy        518 FGGKAKKI--PSGASFRGDINVLLVGDPGTSKSQLLQYVHKIAPRGIYTSGRGSSAVGLT 575
              |||    |                  || :||:||  |||||||||||||||:  :  |||  ||||:||||||||
Db        545 FGGTRKDFSHTGRGKFRAEINILLCGDPGTSKSQLLQYVYNLVPRGQYTSGKGSSAVGLT 604

Qy        576 AYVTKDPETRETVLESGALVLSDRGICCIDEFDKMSDNARSMLHEVMEQQTVSVAKGGII 635
              ||| ||||||: ||::||||||  |||||||||||:::  |||:|||||||||:|:||  |||
Db        605 AYVMKDPETRQLVLQTGALVLSDNGICCIDEFDKMNESTRSVLHEVMEQQTLSIAKAGII 664

Qy        636 ASLNARTSVLACANPSGSRYNARLSVIDNIQLPPTLLSRFDLIYLMLDKPDEQNDRRLAR 695
               ||||||||| |||     |::|   : :  |:|||||  ||||||||:||||    ||   |||||
Db        665 CQLNARTSVLAAANPIESQWNPKKTTIENIQLPHTLLSRFDLIFLMLDPQDEAYDRRLAH 724

Qy        696 HLVALHYENYEVSKQDALDLQTLTAYITYARQHVHPTLSDEAAEDLINGYVEMRQKGNFP 755
              |||||:|::  |  ::::  ||:     |     ||   ||   :  | ||:||:||  ||  ||:||:
Db        725 HLVALYYQSEEQAEEELLDMAVLKDYIAYAHSTIMPRLSEEASQALIEAYVDMRK----I 780

Qy        756 GSSKKVITATPRQLESMIRISEALARMRFSEVVEKVDAAEAVRLLDVALQQSATDHATGT 815
              |||:   :::|  ||||||:||::||  |::|    |    |:   ||   ||     ||:|||||  ||
Db        781 GSSRGMVSAYPRQLESLIRLAEAHAKVRLSNKVEAIDVEEAKRLHREALKQSATDPRTGI 840

Qy        816 IDMDLITTGVSASERIRRANLLAALRELIADKISPGSSSGLKTSQLLEDIRSQSSVDVSL 875
              :|: ::|||:||:  | |:   |    ||::||    :|  |  :   ||   ||  ||||  ||  : ::
Db        841 VDISILTTGMSATSRKRKEELAEALKKLI---LSKGKTPALKYQQLFEDIRGQSDIAITK 897

Qy        876 QDIKNALGSLQGEGFLTVHGDIVK 899
                 : || :|   : |||| |   |:
Db        898 DMFEEALRALADDDFLTVTGKTVR 921
```

RESULT 10
ADO19848
ID   ADO19848 standard; protein; 923 AA.
XX
AC   ADO19848;
XX
DT   12-AUG-2004 (first entry)
XX
DE   Human PRO polypeptide #386.
XX
KW   Human; PRO; immune related disorder; systemic lupus erythematosus;

```
KW   rheumatoid arthritis; osteoarthritis; juvenile chronic arthritis;
KW   systemic sclerosis; Sjogren's syndrome; vasculitis; sarcoidosis;
KW   autoimmune haemolytic anaemia; autoimmune thrombocytopenia; thyroiditis;
KW   diabetes mellitus; renal disease; demyelinating disease;
KW   central nervous system; peripheral nervous system;
KW   demyelinating polyneuropathy; Guillain-Barre syndrome;
KW   chronic inflammatory demyelinating polyneuropathy.
XX
OS   Homo sapiens.
XX
PN   WO2004043361-A2.
XX
PD   27-MAY-2004.
XX
PF   06-NOV-2003; 2003WO-US035268.
XX
PR   08-NOV-2002; 2002US-0425235P.
XX
PA   (GETH ) GENENTECH INC.
XX
PI   Fong S,  Dennis K,  Clark H,  Chiu H,  Schoenfeld J,  Williams PM;
PI   Wood WI,  Wu TD;
XX
DR   WPI; 2004-420067/39.
DR   N-PSDB; ADO19847.
XX
PT   Novel PRO polypeptide e.g., PRO69614, PRO71106, or PRO86388 useful for
PT   treating an immune related disorder such as systemic lupus erythematosus,
PT   rheumatoid arthritis, osteoarthritis, juvenile chronic arthritis or
PT   spondyloarthropathy.
XX
PS   Claim 7; SEQ ID NO 772; 1731pp; English.
XX
CC   The invention relates to human PRO polypeptides and the polynucleotides
CC   encoding them. The polypeptides and polynucleotides are useful for
CC   treating and diagnosing immune related disorders in mammals. The immune
CC   related disorders include systemic lupus erythematosus, rheumatoid
CC   arthritis, osteoarthritis, juvenile chronic arthritis, systemic
CC   sclerosis, Sjogren's syndrome, vasculitis, sarcoidosis, autoimmune
CC   haemolytic anaemia, autoimmune thrombocytopenia, thyroiditis, diabetes
CC   mellitus, immune-mediated renal disease, demyelinating diseases of the
CC   central or peripheral nervous system, demyelinating polyneuropathy,
CC   Guillain-Barre syndrome and chronic inflammatory demyelinating
CC   polyneuropathy. This sequence represents a human PRO polypeptide of the
CC   invention.
XX
SQ   Sequence 923 AA;

  Query Match             39.4%;  Score 1817.5;  DB 8;  Length 923;
  Best Local Similarity   43.7%;  Pred. No. 2.5e-142;
  Matches  404;  Conservative  159;  Mismatches  275;  Indels   86;  Gaps   19;

Qy         18 SQSEGVSTPLP-QVTSPSFDNAASPVA-----GRRAVRQTPTSAVRRR-GRETDSARRRR 70
              |:  | :   | : :||    :|| :      | |  | ||    |    | : | |||
Db         42 SREAGPARACPCRAGTPS--TMSSPASTPSRRGSRRGRATPAQTPRSEDARSSPSQRRRG 99

Qy         71 SRSRSLG------------------NSVYSSPYDAGTPGTP-GTPVATPV-YATPVGTP 109
                | | |                 : ::|||     :    |     |::|: | ||
Db        100 EDSTSTGELQPMPTSPGVDLQSTAAQDVLFSSPPQMHSSAIPLDFDVSSPLTYGTPSSRV 159

Qy        110 MGTP-SFHRGTPQYKQRSELGSQGKPLHRRRRSQSREPGHRSPSREPSADGRPSESAEPD 168
               ||| |  ||||  :|| :||| | |      :|                 || :|
Db        160 EGTPRSGVRGTP-VRQRPDLGSAQKGLQVDLQS---------------DGAAAEDIVAS 202

Qy        169 DTLGGEYAYVWGTNVNIPDVLRAIRRFLHNY--------RSSAHDLNSK-YIQIIEETVE 219
               :    |:   :|||:||:        :|||   :          :   |:     |:| : |
Db        203 EQSLGQKLVIWGTDVNVAACKENFQRFLQRFIDPLAKEEENVGIDITEPLYMQRLGEINV 262

Qy        220 REEDTLNIDMSDIYDHDPDLYAKIVRYPLDIIPLLDTECQEVATSLLP--TFEKHIEARP 277
                |  ||::    |    | :|| :::  || ::||   |     |:       |      |  |: ||
Db        263 IGEQFLNVNCEHIKSFDKNLYRQLISYPQEVIPTFDMAVNEIFFDRYPDSILEHQIQVRP 322

Qy        278 FNLKASVHMRELNPSDIDKLVSVKGMVIRCSSIIPEIKGAFFKCLVCGHSPPLVTVVKGR 337
              ||    : :|| ||| |||:|::: ||||| | :|||:: |||:| || |:    | : :||
Db        323 FNALKTKNMRNLNPEDIDQLITISGMVIRTSQLIPEMQEAFFQCQVCAHT-TRVEMDRGR 381

Qy        338 VEEPTRCEKPECAARNAMSLIHNRCTFANKQIVRLQETPDAIPEGETPHTVSMCLYNTMV 397
              : ||: | :   |   ::|:|||||  |::||:::|||:|: :| |:||||| :  :| :|
Db        382 IAEPSVCGR--CHTTHSMALIHNRSLFSDKQMIKLQESPEDMPAGQTPHTVILFAHNDLV 439

Qy        398 DAVKPGDRIEVTGVFKAMAVRVGPNQRTLRALYKTYIDCVHVKKSDRGRLQTEDPMEMDK 457
              | |:||||: |||::::|:  :|| |       ::::|||:||  :| :|:|   ||
Db        440 DKVQPGDRVNVTGIYRAVPIRVNPRVSNVKSVYKTHIDVIHYRKTDAKRLH--------- 490

Qy        458 ENDMYAGYHESDTSEAANEAKIQKLKELSKLPGIYDRLSRSLAPSIWELEDIKKGLLCQL 517
                      |    |     :  :| :::  ||||:  |  ||:||: :||||:| ||||:| ||
Db        491 ------GLDEEAEQKLFSEKRVELLKELSRKPDIYERLASALAPSIYEHEDIKKGILLQL 544

Qy        518 FGGKAKKI--PSGASFRGDINVLLVGDPGTSKSQLLQYVHKIAPRGIYTSGRGSSAVGLT 575
              ||| |            || :||:|| ||||||||||||: : ||| ||||:||||||||
Db        545 FGGTRKDFSHTGRGKFRAEINILLCGDPGTSKSQLLQYVYNLVPRGQYTSGKGSSAVGLT 604

Qy        576 AYVTKDPETRETVLESGALVLSDRGICCIDEFDKMSDNARSMLHEVMEQQTVSVAKGGII 635
              ||| ||||||: ||::||||||| |||||||||||:::: ||:||||||||||:|:|| |||
Db        605 AYVMKDPETRQLVLQTGALVLSDNGICCIDEFDKMNESTRSVLHEVMEQQTLSIAKAGII 664

Qy        636 ASLNARTSVLACANPSGSRYNARLSVIDNIQLPPTLLSRFDLIYLMLDKPDEQNDRRLAR 695
                |||||||| |||  |::| : : |:||||| ||||||||||:||||  ||  |||||
Db        665 CQLNARTSVLAAANPIESQWNPKKTTIENIQLPHTLLSRFDLIFLMLDPQDEAYDRRLAH 724

Qy        696 HLVALHYENYEVSKQDALDLQTLTAYITYARQHVHPTLSDEAAEDLINGYVEMRQKGNFP 755
              |||||:|:: | :::: ||:   |  || ||     :  | ||:||:: ||  ||:||:
Db        725 HLVALYYQSEEQAEEELLDMAVLKDYIAYAHSTIMPRLSEEASQALIEAYVDMRK----I 780

Qy        756 GSSKKVITATPRQLESMIRISEALARMRFSEVVEKVDAAEAVRLLDVALQQSATDHATGT 815
              |||: :::| ||||||:||::|| |::| |   || :|  || ||    ||:||||| ||
Db        781 GSSRGMVSAYPRQLESLIRLAEAHAKVRLSNKVEAIDVEEAKRLHREALKQSATDPRTGI 840

Qy        816 IDMDLITTGVSASERIRRANLLAALRELIADKISPGSSSGLKTSQLLEDIRSQSSVDVSL 875
              :|: ::|||:||: | |:  |  ||::||   :| | :  ||  || ||| || : ::
Db        841 VDISILTTGMSATSRKRKEELAEALKKLI---LSKGKTPALKYQQLFEDIRGQSDIAITK 897

Qy        876 QDIKNALGSLQGEGFLTVHGDIVK 899
                  : || :|  : |||| |  |:
Db        898 DMFEEALRALADDDFLTVTGKTVR 921


RESULT 11
ADU06294
ID   ADU06294 standard; protein; 924 AA.
XX
AC   ADU06294;
XX
DT   27-JAN-2005  (first entry)
XX
DE   Novel bronchial cancer-associated human protein SeqID518.
XX
KW   bronchial cancer; cytostatic; tumour-associated protein;
KW   cancer detection; metastasis; tumour; human.
XX
OS   Homo sapiens.
XX
PN   DE10316701-A1.
XX
PD   04-NOV-2004.
XX
PF   09-APR-2003; 2003DE-01016701.
XX
PR   09-APR-2003; 2003DE-01016701.
XX
PA   (HINZ/) HINZMANN B.
PA   (HERM/) HERMANN K.
PA   (CAST/) HEIDEN CASTANOS-VELEZ E.
XX
PI   Mennerich D,  Bruemmendorf T,  Heiden E,  Hermann K,  Kinnemann H;
PI   Li X,  Roepcke S,  Staub E,  Hinzmann B,  Rosenthal A,  Pilarsky C;
```

```
XX
DR   WPI; 2004-786403/78.
DR   N-PSDB; ADU05807.
XX
PT   New nucleic acid, and derived proteins, useful for diagnosis of bronchial
PT   cancer and in screening for therapeutic and diagnostic agents.
XX
PS   Claim 2; SEQ ID NO 518; 1381pp; German.
XX
CC   This invention relates to a novel isolated nucleic acid associated with
CC   bronchial cancer comprising 489 defined sequences given in the
CC   specification. The invention may be useful for the production of
CC   compounds with a cytostatic activity through the inhibition of expression
CC   or activity of tumour-associated proteins. The novel DNA sequences and
CC   the proteins/peptides encoded by them are used for detecting bronchial
CC   cancer or determining the risk of developing it and to screen for
CC   specific binding partners of the DNA or protein sequences, where the
CC   binding partners are potentially useful as agents for treating or
CC   diagnosing bronchial cancer. The DNA or protein sequences can also be
CC   used for prognosis, detection of metastases and for secondary treatment
CC   (of tumours that have been stabilised or are no longer detectable).
CC   Detecting abnormal expression of the DNA sequences provides early
CC   diagnosis of bronchial cancers. The present sequence is that of a protein
CC   encoded by a novel bronchial cancer-associated human gene sequence of the
CC   invention.
XX
SQ   Sequence 924 AA;

  Query Match             39.4%;  Score 1817.5;  DB 8;  Length 924;
  Best Local Similarity   43.7%;  Pred. No. 2.5e-142;
  Matches  404;  Conservative  159;  Mismatches  275;  Indels   86;  Gaps   19;

Qy         18 SQSEGVSTPLP-QVTSPSFDNAASPVA-----GRRAVRQTPTSAVRRR-GRETDSARRRR 70
Db         42 SREAGPARACPCRAGTPS--TMSSPASTPSRRGSRRGRATPAQTPRSEDARSSPSQRRRG 99

Qy         71 SRSRSLG------------------NSVYSSPYDAGTPGTP-GTPVATPV-YATPVGTP 109
Db        100 EDSTSTGELQPMPTSPGVDLQSTAAQDVLFSSPPQMHSSAIPLDFDVSSPLTYGTPSSRV 159

Qy        110 MGTP-SFHRGTPQYKQRSELGSQGKPLHRRRRSQSREPGHRSPSREPSADGRPSESAEPD 168
Db        160 EGTPRSGVRGTP-VRQRPDLGSAQKGLQVDLQS---------------DGAAAEDIVAS 202

Qy        169 DTLGGEYAYVWGTNVNIPDVLRAIRRFLHNY--------RSSAHDLNSK-YIQIIEETVE 219
Db        203 EQSLGQKLVIWGTDVNVAACKENFQRFLQRFIDPLAKEEENVGIDITEPLYMQRLGEINV 262

Qy        220 REEDTLNIDMSDIYDHDPDLYAKIVRYPLDIIPLLDTECQEVATSLLP--TFEKHIEARP 277
Db        263 IGEQFLNVNCEHIKSFDKNLYRQLISYPQEVIPTFDMAVNEIFFDRYPDSILEHQIQVRP 322

Qy        278 FNLKASVHMRELNPSDIDKLVSVKGMVIRCSSIIPEIKGAFFKCLVCGHSPPLVTVVKGR 337
Db        323 FNALKTKNMRNLNPEDIDQLITISGMVIRTSQLIPEMQEAFFQCQVCAHT-TRVEMDRGR 381

Qy        338 VEEPTRCEKPECAARNAMSLIHNRCTFANKQIVRLQETPDAIPEGETPHTVSMCLYNTMV 397
Db        382 IAEPSVCGR--CHTTHSMALIHNRSLFSDKQMIKLQESPEDMPAGQTPHTVILFAHNDLV 439

Qy        398 DAVKPGDRIEVTGVFKAMAVRVGPNQRTLRALYKTYIDCVHVKKSDRGRLQTEDPMEMDK 457
Db        440 DKVQPGDRVNVTGIYRAVPIRVNPRVSNVKSVYKTHIDVIHYRKTDAKRLH--------- 490

Qy        458 ENDMYAGYHESDTSEAANEAKIQKLKELSKLPGIYDRLSRSLAPSIWELEDIKKGLLCQL 517
Db        491 ------GLDEEAEQKLFSEKRVELLKELSRKPDIYERLASALAPSIYEHEDIKKGILLQL 544

Qy        518 FGGKAKKI--PSGASFRGDINVLLVGDPGTSKSQLLQYVHKIAPRGIYTSGRGSSAVGLT 575
Db        545 FGGTRKDFSHTGRGKFRAEINILLCGDPGTSKSQLLQYVYNLVPRGQYTSGKGSSAVGLT 604

Qy        576 AYVTKDPETRETVLESGALVLSDRGICCIDEFDKMSDNARSMLHEVMEQQTVSVAKGGII 635
Db        605 AYVMKDPETRQLVLQTGALVLSDNGICCIDEFDKMNESTRSVLHEVMEQQTLSIAKAGII 664

Qy        636 ASLNARTSVLACANPSGSRYNARLSVIDNIQLPPTLLSRFDLIYLMLDKPDEQNDRRLAR 695
Db        665 CQLNARTSVLAAANPIESQWNPKKTTIENIQLPHTLLSRFDLIFLMLDPQDEAYDRRLAH 724

Qy        696 HLVALHYENYEVSKQDALDLQTLTAYITYARQHVHPTLSDEAAEDLINGYVEMRQKGNFP 755
Db        725 HLVALYYQSEEQAEEELLDMAVLKDYIAYAHSTIMPRLSEEASQALIEAYVDMRK----I 780

Qy        756 GSSKKVITATPRQLESMIRISEALARMRFSEVVEKVDAAEAVRLLDVALQQSATDHATGT 815
Db        781 GSSRGMVSAYPRQLESLIRLAEAHAKVRLSNKVEAIDVEEAKRLHREALKQSATDPRTGI 840

Qy        816 IDMDLITTGVSASERIRRANLLAALRELIADKISPGSSSGLKTSQLLEDIRSQSSVDVSL 875
Db        841 VDISILTTGMSATSRKRKEELAEALKKLI---LSKGKTPALKYQQLFEDIRGQSDIAITK 897

Qy        876 QDIKNALGSLQGEGFLTVHGDIVK 899
Db        898 DMFEEALRALADDDFLTVTGKTVR 921

RESULT 12
ADR69821
ID   ADR69821 standard; protein; 863 AA.
XX
AC   ADR69821;
XX
DT   02-DEC-2004  (first entry)
XX
DE   Human minichromosome maintenance-4 (Mcm4) protein.
XX
KW   synthesis-period integration effect; cell cycle; phosphorylation;
KW   minichromosome maintenance-4; Mcm4; Chk-1; cancer; Cdk-2; cytostatic;
KW   human.
XX
OS   Homo sapiens.
XX
PN   JP2004248594-A.
XX
PD   09-SEP-2004.
XX
PF   20-FEB-2003; 2003JP-00043174.
XX
PR   20-FEB-2003; 2003JP-00043174.
XX
PA   (MITU ) MITSUBISHI CHEM CORP.
XX
DR   WPI; 2004-656246/64.
XX
PT   Screening synthesis-period integration effect releasing substance of
PT   cell, useful in treating cancer, by culturing cell with test substance
PT   and selecting substance suppressing phosphorylation of minichromosome
PT   maintenance-4 protein.
XX
PS   Example 1; SEQ ID NO 1; 30pp; Japanese.
XX
CC   The invention relates to a novel method for screening a substance capable
CC   of releasing the synthesis-period integration effect of a cell. The
CC   method involves culturing the cell in the presence of the test substance,
CC   where the cell cycle has been stopped by the synthesis period, and
CC   selecting the substance, which has the capability to suppress
CC   phosphorylation of minichromosome maintenance-4 (Mcm4) protein, without
CC   inhibiting the function of the Chk-1 protein in the cell. The invention
CC   further comprises: a substance obtained by the method; a therapeutic drug
CC   for treating cancer, comprising the substance obtained by the method, and
CC   a DNA reproduction inhibitor; a medical agent for increasing the
CC   therapeutic effect of treating the cancer by DNA reproduction inhibitor,
CC   comprising the substance obtained by the method as an active ingredient;
CC   and a medical agent for releasing the synthesis period integration effect
CC   of a cell, comprising the inhibitor of the Cdk-2 protein as an active
```

```
CC   ingredient. The therapeutic drug has cytostatic activity. The method is
CC   useful for screening a substance capable of releasing the synthesis-
CC   period integration effect of a cell. The substance obtained by the method
CC   is useful as a therapeutic drug with respect to cancer. The method
CC   enables the obtaining of a substance having high therapeutic effect and
CC   lesser side effects when used in treating cancer. This sequence
CC   represents a human minichromosome maintenance-4 (Mcm4) protein of the
CC   invention.
XX
SQ   Sequence 863 AA;

  Query Match             39.4%;  Score 1814.5;  DB 8;  Length 863;
  Best Local Similarity   44.2%;  Pred. No. 4e-142;
  Matches  401;  Conservative  157;  Mismatches  266;  Indels   83;  Gaps   18;

Qy         30 VTSPSFDNAASPV-AGRRAVRQTPTSAVRRR-GRETDSARRRRSRSRSLG---------- 77
              ::||     |::|    |  |   |  ||      |     | : |  |||    | | |
Db          1 MSSP----ASTPSRRGSRRGRATPAQTPRSEDARSSPSQRRRGEDSTSTGELQPMPTSPG 56

Qy         78 ---------NSVYSSPYDAGTPGTP-GTPVATPV-YATPVGTPMGTP-SFHRGTPQYKQR 125
                       : ::|||     :     |     |::|: | ||      ||| |   |||| :||
Db         57 VDLQSTAAQDVLFSSPPQMHSSAIPLDFDVSSPLTYGTPSSRVEGTPRSGVRGTP-VRQR 115

Qy        126 SELGSQGKPLHRRRRSQSREPGHRSPSREPSADGRPSESAEPDDTLGGEYAYVWGTNVNI 185
                :|||  | |     :|                       || :|      :     |:   :|||:||:
Db        116 PDLGSAQKGLQVDLQS---------------DGAAAEDIVASEQSLGQKLVIWGTDVNV 159

Qy        186 PDVLRAIRRFLHNY--------RSSAHDLNSK-YIQIIEETVEREEDTLNIDMSDIYDHD 236
                     :|||    :           :      |:      |:| :  |       |  ||::      |
Db        160 AACKENFQRFLQRFIDPLAKEEENVGIDITEPLYMQRLGEINVIGEQFLNVNCEHIKSFD 219

Qy        237 PDLYAKIVRYPLDIIPLLDTECQEVATSLLP--TFEKHIEARPFNLKASVHMRELNPSDI 294
              :|| ::: || ::||   |     |:           |     |  |: ||||    : :|| ||| ||
Db        220 KNLYRQLISYPQEVIPTFDMAVNEIFFDRYPDSILEHQIQVRPFNALKTKNMRNLNPEDI 279

Qy        295 DKLVSVKGMVIRCSSIIPEIKGAFFKCLVCGHSPPLVTVVKGRVEEPTRCEKPECAARNA 354
              |:|::: ||||| | :|||:: |||:| || |:     | : :||: ||: |  :  |     ::
Db        280 DQLITISGMVIRTSQLIPEMQEAFFQCQVCAHT-TRVEMDRGRIAEPSVCGR--CHTTHS 336

Qy        355 MSLIHNRCTFANKQIVRLQETPDAIPEGETPHTVSMCLYNTMVDAVKPGDRIEVTGVFKA 414
              |:||||| |::||:::|||:|: :| |:||||| :   :| :|| |:||||: |||:::|
Db        337 MALIHNRSLFSDKQMIKLQESPEDMPAGQTPHTVILFAHNDLVDKVQPGDRVNVTGIYRA 396

Qy        415 MAVRVGPNQRTLRALYKTYIDCVHVKKSDRGRLQTEDPMEMDKENDMYAGYHESDTSEAA 474
              : :|| |     ::::||:|| :| :|:| ||                          |  |       :
Db        397 VPIRVNPRVSNVKSVYKTHIDVIHYRKTDAKRLH--------------GLDEEAEQKLF 441

Qy        475 NEAKIQKLKELSKLPGIYDRLSRSLAPSIWELEDIKKGLLCQLFGGKAKKI--PSGASFR 532
              :| ::: |||||: | ||:||: :|||||:| ||||||:| |||||           ||
Db        442 SEKRVELLKELSRKPDIYERLASALAPSIYEHEDIKKGILLQLFGGTRKDFSHTGRGKFR 501

Qy        533 GDINVLLVGDPGTSKSQLLQYVHKIAPRGIYTSGRGSSAVGLTAYVTKDPETRETVLESG 592
               :||:|| |||||||||||||: : ||| ||||:||||||||||| ||||||: ||::|
Db        502 AEINILLCGDPGTSKSQLLQYVYNLVPRGQYTSGKGSSAVGLTAYVMKDPETRQLVLQTG 561

Qy        593 ALVLSDRGICCIDEFDKMSDNARSMLHEVMEQQTVSVAKGGIIASLNARTSVLACANPSG 652
              ||||| ||||||||||:::: ||:|||||||||:|:|| ||| |||||||| |||
Db        562 ALVLSDNGICCIDEFDKMNESTRSVLHEVMEQQTLSIAKAGIICQLNARTSVLAAANPIE 621

Qy        653 SRYNARLSVIDNIQLPPTLLSRFDLIYLMLDKPDEQNDRRLARHLVALHYENYEVSKQDA 712
              |::| : : |:||||| ||||||||||:||||  ||  |||||  ||||:|:: | ::::
Db        622 SQWNPKKTTIENIQLPHTLLSRFDLIFLMLDPQDEAYDRRLAHHLVALYYQSEEQAEEEL 681

Qy        713 LDLQTLTAYITYARQHVHPTLSDEAAEDLINGYVEMRQKGNFPGSSKKVITATPRQLESM 772
              ||:  |  ||  ||    :  | ||:||::  ||  ||:||:      |||: :::| ||||||:
Db        682 LDMAVLKDYIAYAHSTIMPRLSEEASQALIEAYVDMRK----IGSSRGMVSAYPRQLESL 737

Qy        773 IRISEALARMRFSEVVEKVDAAEAVRLLDVALQQSATDHATGTIDMDLITTGVSASERIR 832
              ||::|| |::| |  || :|  || ||     ||:|||||  || :|: ::|||:||: | |
Db        738 IRLAEAHAKVRLSNKVEAIDVEEAKRLHREALKQSATDPRTGIVDISILTTGMSATSRKR 797

Qy        833 RANLLAALRELIADKISPGSSSGLKTSQLLEDIRSQSSVDVSLQDIKNALGSLQGEGFLT 892
              :  |  ||::||    :| |  :  ||   || |||| || : ::      : || :|  : |||
Db        798 KEELAEALKKLI---LSKGKTPALKYQQLFEDIRGQSDIAITKDMFEEALRALADDDFLT 854

Qy        893 VHGDIVK 899
               | |  |:
Db        855 VTGKTVR 861

RESULT 13
AAB95203
ID   AAB95203 standard; protein; 863 AA.
XX
AC   AAB95203;
XX
DT   26-JUN-2001  (first entry)
XX
DE   Human protein sequence SEQ ID NO:17301.
XX
KW   Human; primer; detection; diagnosis; antisense therapy; gene therapy.
XX
OS   Homo sapiens.
XX
PN   EP1074617-A2.
XX
PD   07-FEB-2001.
XX
PF   28-JUL-2000; 2000EP-00116126.
XX
PR   29-JUL-1999;   99JP-00248036.
PR   27-AUG-1999;   99JP-00300253.
PR   11-JAN-2000; 2000JP-00118776.
PR   02-MAY-2000; 2000JP-00183767.
PR   09-JUN-2000; 2000JP-00241899.
XX
PA   (HELI-) HELIX RES INST.
XX
PI   Ota T,  Isogai T,  Nishikawa T,  Hayashi K,  Saito K,  Yamamoto J;
PI   Ishii S,  Sugiyama T,  Wakamatsu A,  Nagai K,  Otsuki T;
XX
DR   WPI; 2001-318749/34.
XX
PT   Primer sets for synthesizing polynucleotides, particularly the 5602 full-
PT   length cDNAs defined in the specification, and for the detection and/or
PT   diagnosis of the abnormality of the proteins encoded by the full-length
PT   cDNAs.
XX
PS   Claim 8; SEQ ID NO 17301; 2537pp + Sequence Listing; English.
XX
CC   The present invention describes primer sets for synthesising 5602 full-
CC   length cDNAs defined in the specification. Where a primer set comprises:
CC   (a) an oligo-dT primer and an oligonucleotide complementary to the
CC   complementary strand of a polynucleotide which comprises one of the 5602
CC   nucleotide sequences defined in the specification, where the
CC   oligonucleotide comprises at least 15 nucleotides; or (b) a combination
CC   of an oligonucleotide comprising a sequence complementary to the
CC   complementary strand of a polynucleotide which comprises a 5'-end
CC   sequence and an oligonucleotide comprising a sequence complementary to a
CC   polynucleotide which comprises a 3'-end sequence, where the
CC   oligonucleotide comprises at least 15 nucleotides and the combination of
CC   the 5'-end sequence/3'-end sequence is selected from those defined in the
CC   specification. The primer sets can be used in antisense therapy and in
CC   gene therapy. The primers are useful for synthesising polynucleotides,
CC   particularly full-length cDNAs. The primers are also useful for the
CC   detection and/or diagnosis of the abnormality of the proteins encoded by
CC   the full-length cDNAs. The primers allow obtaining of the full-length
CC   cDNAs easily without any specialised methods. AAH03166 to AAH13628 and
CC   AAH13633 to AAH18742 represent human cDNA sequences; AAB92446 to AAB95893
CC   represent human amino acid sequences; and AAH13629 to AAH13632 represent
CC   oligonucleotides, all of which are used in the exemplification of the
CC   present invention
XX
SQ   Sequence 863 AA;

  Query Match             39.3%;  Score 1810.5;  DB 4;  Length 863;
```

```
  Best Local Similarity   44.1%;  Pred. No. 8.6e-142;
  Matches  400;  Conservative  158;  Mismatches  266;  Indels   83;  Gaps   18;

Qy         30 VTSPSFDNAASPV-AGRRAVRQTPTSAVRRR-GRETDSARRRRSRSRSLG---------- 77
Db          1 MSSP----ASTPSRRGSRRGRATPAQTPRSEDARSSPSQRRRGEDSTSTGELQPMPTSPG 56

Qy         78 ---------NSVYSSPYDAGTPGTP-GTPVATPV-YATPVGTPMGTP-SFHRGTPQYKQR 125
Db         57 VDLQSPAAQDVLFSSPPQMHSSAIPLDFDVSSPLTYGTPSSRVEGTPRSGVRGTP-VRQR 115

Qy        126 SELGSQGKPLHRRRRSQSREPGHRSPSREPSADGRPSESAEPDDTLGGEYAYVWGTNVNI 185
Db        116 PDLGSAQKGLQVDLQS----------------DGAAAEDIVASEQSLGQKLVIWGTDVNV 159

Qy        186 PDVLRAIRRFLHNY--------RSSAHDLNSK-YIQIIEETVEREEDTLNIDMSDIYDHD 236
Db        160 AACKENFQRFLQRFIDPLAKEEENVGIDITEPLYMQRLGEINVIGEPFLNVNCEHIKSFD 219

Qy        237 PDLYAKIVRYPLDIIPLLDTECQEVATSLLP--TFEKHIEARPFNLKASVHMRELNPSDI 294
Db        220 KNLYRQLISYPQEVIPTFDMAVNEIFFDRYPDSILEHQIQVRPFNALKTKNMRNLNPEDI 279

Qy        295 DKLVSVKGMVIRCSSIIPEIKGAFFKCLVCGHSPPLVTVVKGRVEEPTRCEKPECAARNA 354
Db        280 DQLITISGMVIRTSQLIPEMQEAFFQCQVCAHT-TRVEMDRGRIAEPSVCGR--CHTTHS 336

Qy        355 MSLIHNRCTFANKQIVRLQETPDAIPEGETPHTVSMCLYNTMVDAVKPGDRIEVTGVFKA 414
Db        337 MALIHNRSLFSDKQMIKLQESPEDMPAGQTPHTVILFAHNDLVDKVQPGDRVNVTGIYRA 396

Qy        415 MAVRVGPNQRTLRALYKTYIDCVHVKKSDRGRLQTEDPMEMDKENDMYAGYHESDTSEAA 474
Db        397 VPIRVNPRVSNVKSVYKTHIDVIHYRKTDAKRLH---------------GLDEEAEQKLF 441

Qy        475 NEAKIQKLKELSKLPGIYDRLSRSLAPSIWELEDIKKGLLCQLFGGKAKKI--PSGASFR 532
Db        442 SEKRVELLKELSRKPDIYERLASALAPSIYEHEDIKKGILLQLFGGTRKDFSHTGRGKFR 501

Qy        533 GDINVLLVGDPGTSKSQLLQYVHKIAPRGIYTSGRGSSAVGLTAYVTKDPETRETVLESG 592
Db        502 AEINILLCGDPGTSKSQLLQYVYNLVPRGQYTSGKGSSAVGLTAYVMKDPETRQLVLQTG 561

Qy        593 ALVLSDRGICCIDEFDKMSDNARSMLHEVMEQQTVSVAKGGIIASLNARTSVLACANPSG 652
Db        562 ALVLSDNGICCIDEFDRMNESTRSVLHEVMEQQTLSIAKAGIICQLNARTSVLAAANPIE 621

Qy        653 SRYNARLSVIDNIQLPPTLLSRFDLIYLMLDKPDEQNDRRLARHLVALHYENYEVSKQDA 712
Db        622 SQWNPKKTTIENIQLPHTLLSRFDLIFLMLDPQDEAYDRRLAHHLVALYYQSEEQAEEEL 681

Qy        713 LDLQTLTAYITYARQHVHPTLSDEAAEDLINGYVEMRQKGNFPGSSKKVITATPRQLESM 772
Db        682 LDMAVLKDYIAYAHSTIMPRLSEEASQALIEAYVDMRK----IGSSRGMVSAYPRQLESL 737

Qy        773 IRISEALARMRFSEVVEKVDAAEAVRLLDVALQQSATDHATGTIDMDLITTGVSASERIR 832
Db        738 IRLAEAHAKVRLSNKVEAIDVEEAKRLHREALKQSATDPRTGIVDISILTTGMSATSRKR 797

Qy        833 RANLLAALRELIADKISPGSSSGLKTSQLLEDIRSQSSVDVSLQDIKNALGSLQGEGFLT 892
Db        798 KEELAEALKKLI---LSKGKTPALKYQQLFEDIRGQSDIAITKDMFEEALRALADDDFLT 854

Qy        893 VHGDIVK 899
Db        855 VTGKTVR 861
```

```
RESULT 14
ABM80987
ID   ABM80987 standard; protein; 863 AA.
XX
AC   ABM80987;
XX
DT   18-NOV-2004  (first entry)
XX
DE   Tumour-associated antigenic target (TAT) polypeptide PRO81673, SEQ:2545.
XX
KW   Tumour-associated antigenic target; TAT; human; overexpression; cancer;
KW   tumour; diagnosis; cell proliferative disorder; breast cancer;
KW   colorectal cancer; lung cancer; ovarian cancer; liver cancer;
KW   central nervous system cancer; bladder cancer; pancreatic cancer;
KW   cervical cancer; melanoma; leukaemia; hybridisation probe;
KW   chromosome identification; chromosome mapping; gene mapping;
KW   gene therapy; cytostatic.
XX
OS   Homo sapiens.
XX
PN   WO2004030615-A2.
XX
PD   15-APR-2004.
XX
PF   29-SEP-2003; 2003WO-US028547.
XX
PR   02-OCT-2002; 2002US-0414971P.
XX
PA   (GETH ) GENENTECH INC.
XX
PI   Wu TD,  Zhang Z,  Zhou Y;
XX
DR   WPI; 2004-347921/32.
DR   N-PSDB; ACN38801.
XX
PT   New tumor-associated antigenic target polypeptides and nucleic acids,
PT   useful in preparing a medicament for treating or detecting a
PT   proliferative disorder, e.g. breast, lung, colorectal, ovarian or
PT   prostate cancer or tumor.
XX
PS   Claim 12; SEQ ID NO 2545; 7273pp; English.
XX
CC   The invention relates to human tumour-associated antigenic target (TAT)
CC   polypeptides, and their related nucleic acids. The TAT polypeptides are
CC   overexpressed in cancer tissues compared to normal tissues, and may thus
CC   serve as effective targets for the diagnosis and treatment of cancer in
CC   mammals. The invention also relates to nucleic acid and polypeptide
CC   sequences at least 80% identical to the TAT nucleic acids and
CC   polypeptides; expression vectors and host cells comprising a TAT nucleic
CC   acid; an antibody specific for a TAT polypeptide; a peptide or organic
CC   molecule which binds to a TAT polypeptide; fusion proteins comprising a
CC   TAT polypeptide; and methods and compositions for the treatment or
CC   diagnosis of cancer in mammals. TAT polypeptides, nucleic acids,
CC   antibodies, antagonists, binding molecules and compositions are useful
CC   for diagnosing or treating a cell proliferative disorder associated with
CC   increased TAT expression, particularly cancers such as breast cancer,
CC   colorectal cancer, lung cancer, ovarian cancer, liver cancer, bladder
CC   cancer, pancreatic cancer, cervical cancer, cancers of the central
CC   nervous system, melanoma and leukaemia. TAT nucleic acids may further be
CC   used as hybridisation probes, in chromosome and gene mapping, in
CC   chromosome identification and in gene therapy. The present sequence
CC   represents a TAT polypeptide of the invention
XX
SQ   Sequence 863 AA;

  Query Match             39.3%;  Score 1810.5;  DB 8;  Length 863;
  Best Local Similarity   44.1%;  Pred. No. 8.6e-142;
  Matches  400;  Conservative  158;  Mismatches  266;  Indels   83;  Gaps   18;

Qy         30 VTSPSFDNAASPV-AGRRAVRQTPTSAVRRR-GRETDSARRRRSRSRSLG---------- 77
Db          1 MSSP----ASTPSRRGSRRGRATPAQTPRSEDARSSPSQRRRGEDSTSTGELQPMPTSPG 56

Qy         78 ---------NSVYSSPYDAGTPGTP-GTPVATPV-YATPVGTPMGTP-SFHRGTPQYKQR 125
Db         57 VDLQSPAAQDVLFSSPPQMHSSAIPLDFDVSSPLTYGTPSSRVEGTPRSGVRGTP-VRQR 115

Qy        126 SELGSQGKPLHRRRRSQSREPGHRSPSREPSADGRPSESAEPDDTLGGEYAYVWGTNVNI 185
```

```
                 :|||  | |     :|                    || :|     :   |:   :|||:||:
Db         116 PDLGSAQKGLQVDLQS---------------DGAAAEDIVASEQSLGQKLVIWGTDVNV 159

Qy         186 PDVLRAIRRFLHNY--------RSSAHDLNSK-YIQIIEETVEREEDTLNIDMSDIYDHD 236
                  :|||  :             :    |:    |:| : |      |  ||::    |    |
Db         160 AACKENFQRFLQRFIDPLAKEEENVGIDITEPLYMQRLGEINVIGEPFLNVNCEHIKSFD 219

Qy         237 PDLYAKIVRYPLDIIPLLDTECQEVATSLLP--TFEKHIEARPFNLKASVHMRELNPSDI 294
                :|| ::: || ::||  |     |:      |      |   |: ||||   : :|| ||| ||
Db         220 KNLYRQLISYPQEVIPTFDMAVNEIFFDRYPDSILEHQIQVRPFNALKTKNMRNLNPEDI 279

Qy         295 DKLVSVKGMVIRCSSIIPEIKGAFFKCLVCGHSPPLVTVVKGRVEEPTRCEKPECAARNA 354
               |:|::: ||||| | :|||:: |||:| || |:   | : :||: ||: | :  |    ::
Db         280 DQLITISGMVIRTSQLIPEMQEAFFQCQVCAHT-TRVEMDRGRIAEPSVCGR--CHTTHS 336

Qy         355 MSLIHNRCTFANKQIVRLQETPDAIPEGETPHTVSMCLYNTMVDAVKPGDRIEVTGVFKA 414
               |:||||  |::|:::|||:|: :| |:||||  :   :| :|| |:||||: |||:::|
Db         337 MALIHNRSLFSDKQMIKLQESPEDMPAGQTPHTVILFAHNDLVDKVQPGDRVNVTGIYRA 396

Qy         415 MAVRVGPNQRTLRALYKTYIDCVHVKKSDRGRLQTEDPMEMDKENDMYAGYHESDTSEAA 474
               : :|| |    ::::|||:|| :| :|:| ||                     | |   :
Db         397 VPIRVNPRVSNVKSVYKTHIDVIHYRKTDAKRLH--------------GLDEEAEQKLF 441

Qy         475 NEAKIQKLKELSKLPGIYDRLSRSLAPSIWELEDIKKGLLCQLFGGKAKKI--PSGASFR 532
               :| ::: |||||: | ||:||: :|||||:| ||||||:| |||||  |          ||
Db         442 SEKRVELLKELSRKPDIYERLASALAPSIYEHEDIKKGILLQLFGGTRKDFSHTGRGKFR 501

Qy         533 GDINVLLVGDPGTSKSQLLQYVHKIAPRGIYTSGRGSSAVGLTAYVTKDPETRETVLESG 592
                :||:|| ||||||||||||||: : ||| ||||:||||||||||||| ||||||: ||::|
Db         502 AEINILLCGDPGTSKSQLLQYVYNLVPRGQYTSGKGSSAVGLTAYVMKDPETRQLVLQTG 561

Qy         593 ALVLSDRGICCIDEFDKMSDNARSMLHEVMEQQTVSVAKGGIIASLNARTSVLACANPSG 652
               ||||| |||||||||||:::: ||:||||||||||:|:|| |||  ||||||||| |||
Db         562 ALVLSDNGICCIDEFDKMNESTRSVLHEVMEQQTLSIAKAGIICQLNARTSVLAAANPIE 621

Qy         653 SRYNARLSVIDNIQLPPTLLSRFDLIYLMLDKPDEQNDRRLARHLVALHYENYEVSKQDA 712
               |::| : : |:||||| ||||||||||:|:||  || ||||| |||||:|:: | ::::
Db         622 SQWNPKKTTIENIQLPHTLLSRFDLIFLLLDPQDEAYDRRLAHHLVALYYQSEEQAEEEL 681

Qy         713 LDLQTLTAYITYARQHVHPTLSDEAAEDLINGYVEMRQKGNFPGSSKKVITATPRQLESM 772
               ||:  | || ||    : | ||:||:: || ||:||:       |||: :::| ||||||:
Db         682 LDMAVLKDYIAYAHSTIMPRLSEEASQALIEAYVDMRK----IGSSRGMVSAYPRQLESL 737

Qy         773 IRISEALARMRFSEVVEKVDAAEAVRLLDVALQQSATDHATGTIDMDLITTGVSASERIR 832
               ||:::|| |::| |  ||  :|    || ||     |||:||||  ||  :|: ::|||:|: | |
Db         738 IRLAEAHAKVRLSNKVEAIDVEEAKRLHREALKQSATDPRTGIVDISILTTGMSATSRKR 797

Qy         833 RANLLAALRELIADKISPGSSSGLKTSQLLEDIRSQSSVDVSLQDIKNALGSLQGEGFLT 892
                 :   |  ||::||  :| :  ||  ||    |||:||||  || : :     :  || :|  :  |||
Db         798 KEELAEALKKLI---LSKGKTPALKYQQLFEDIRGQSDIAITKDMFEEALRALADDDFLT 854

Qy         893 VHGDIVK 899
               | |  |:
Db         855 VTGKTVR 861


RESULT 15
ABP73643
ID   ABP73643 standard; protein; 910 AA.
XX
AC   ABP73643;
XX
DT   30-JAN-2003  (first entry)
XX
DE   Candida albicans essential protein SEQ ID NO 7480.
XX
KW   Fungus; yeast; tetracyclin; promoter; GRACE strain; biosynthesis;
KW   signal transduction; DNA replication; cell division; growth;
KW   proliferation; Candida albicans; fungicide; antifungal.
XX
OS   Candida albicans.
XX
PN   WO200253728-A2.
XX
PD   11-JUL-2002.
XX
PF   26-DEC-2001; 2001WO-US049486.
XX
PR   29-DEC-2000; 2000US-0259128P.
PR   20-FEB-2001; 2001US-00792024.
PR   22-AUG-2001; 2001US-0314050P.
XX
PA   (ELIT-) ELITRA PHARM INC.
XX
PI   Roemer T,  Jiang B,  Boone C,  Bussey H,  Ohlsen KL;
XX
DR   WPI; 2002-566694/60.
DR   N-PSDB; ABZ32193.
XX
PT   Constructing strains for identifying gene products as effective targets
PT   for therapeutic intervention, by inactivating in the strain one allele of
PT   a gene and placing other allele of the gene under conditional expression.
XX
PS   Claim 44; SEQ ID NO 7480; 167pp + Sequence Listing; English.
XX
CC   The invention relates to constructing (M1) a strain of diploid fungal
CC   cells in which both alleles of a gene are modified, comprising modifying
CC   one allele by insertion or replacement by a cassette having an
CC   expressible selectable marker and modifying other allele by
CC   recombination, of a promoter replacement fragment with a heterologous
CC   promoter, so that expression of the second allele is regulated by the
CC   promoter. (M1) is useful for constructing a strain of diploid fungal
CC   cells in which both alleles of a gene are modified. The diploid fungal
CC   cells having both alleles modified are useful for identifying a gene that
CC   is essential to the survival or growth of a fungus, a gene that
CC   contributes to the virulence and/or pathogenicity of a fungus, a gene
CC   that contributes to the resistance of a diploid fungus to an antifungal
CC   agent, an antifungal agent that inhibits the growth of a diploid fungus
CC   and for identifying a therapeutic agent for treatment of a mammalian
CC   disease. (M1) is useful for identifying a compound which modulates the
CC   activity of a gene product, preferably enzymatic activity, carbon
CC   compound catabolism, biosynthetic, transporter, transcriptional,
CC   translational, signal transduction, DNA replication and cell division
CC   activity. The method is useful for identifying a compound having the
CC   ability to inhibit growth or proliferation of C. albicans cells and for
CC   treating infection by C. albicans. The present sequence is that of an
CC   essential Candida albicans protein used in the method of the invention.
CC   Note: The sequence data for this patent is not represented in the printed
CC   specification but is based on sequence information supplied to Derwent by
CC   the European Patent Office
XX
SQ   Sequence 910 AA;

  Query Match             38.1%;  Score 1754;  DB 5;  Length 910;
  Best Local Similarity   42.0%;  Pred. No. 5.1e-137;
  Matches  397;  Conservative  155;  Mismatches  259;  Indels  134;  Gaps   20;

Qy           3 NNDALDIGAVSSPYPSQSE---GVSTPLPQVTSPSFDNAASPVAGRRAVRQTPTSAV--- 56
                :|   |||  :|   |||:     :|:||    :|                   ||| :
Db          45 SNPGSDIGNFNSQSQSQSQRRNDISSPLHYTSSAQ----------------PTSDIGGF 87

Qy          57 --RRRGRETDSARRRRSRSRSLGNSVYSSPYDAGTPGTPGTPVATPVYATPVGTPMGTPS 114
                 :|  |   |   |   |    ||        |||
Db          88 DSQRSARVQDVGRIMRRAQRSDVTDSVSSP----------------------------- 117

Qy         115 FHRGTPQYKQRSELGSQGKPLHRRRRSQSREPGHRSPSREPSADGRPSESAEPDDTLGGE 174
                          |||           ||  :|  |   |   :  :     ||       :: |||:|
Db         118 ---------QRS----------RRYFTQGRGNGPSNLNSSTSAQ-FSTDPAEPNDE---P 154

Qy         175 YAYVWGTNVNIPDVLRAIRRFL----HNYRSSA---------HDLNSKYIQIIEETVERE 221
                  :||||:|  :          | ||       :  ||                  |:|   |:  :   :|
Db         155 VRVIWGTNVSIQECSNIFRDFLLSFKYKYRRELEEQAIEPEDHEL--YYVNQLNNIIELG 212

Qy         222 EDTLNIDMSDI--YDHDPDLYAKIVRYPLDIIPLLDTECQEV-------------ATSL 265
                    ||:|   ::  |       || ::: || :|||::|    ::                     | |
Db         213 LTNLNLDAKNLLSYPSTRKLYYQLINYPQEIIPIMDHTIKDCLIQIINDANATTSPAQSK 272
```

```
Qy    266 LPTFEKHI-EARPFNLK-ASVHMRELNPSDIDKLVSVKGMVIRCSSIIPEIKGAFFKCLV 323
Db    273 LDEIETNVYTIRPYNVNLVEKGIRELNPNDIDKLVSVKGLTLRSTSIIPDMKVAFFRCNA 332

Qy    324 CGHSPPLVTVVKGRVEEPTRCEKPECAARNAMSLIHNRCTFANKQIVRLQETPDAIPEGE 383
Db    333 CGHTVG-VEIDRGVISEPTKCPREVCGQTNSMVLIHNRSSFSDKQVIKLQETPDLVPDGQ 391

Qy    384 TPHTVSMCLYNTMVDAVKPGDRIEVTGVFKAMAVRVGPNQRTLRALYKTYIDCVHVKKSD 443
Db    392 TPHSINLCVYDELVDSCRAGDRVEVCGIFRSTPVRANPRQRALKNLYKTYLDIVHVKKID 451

Qy    444 RGRLQTE------DPMEMDKENDMYAGYHESDTSEAANEAKIQKLKELSKLPGIYDRLSR 497
Db    452 KRRLGGDVTTLEHELAEKDQE------VEQVRKITAEEEAKI---KEISERDDLYEILAR 502

Qy    498 SLAPSIWELEDIKKGLLCQLFGGKAKKIPSGASFRGDINVLLVGDPGTSKSQLLQYVHKI 557
Db    503 SLAPSIYEMDDVKKGILLQLFGGTNKTFTKGGRYRGDINILLCGDPSTSKSQILQYVHKI 562

Qy    558 APRGIYTSGRGSSAVGLTAYVTKDPETRETVLESGALVLSDRGICCIDEFDKMSDNARSM 617
Db    563 APRGVYTSGKGSSAVGLTAYITRDIDTKQLVLESGALVLSDGGVCCIDEFDKMSDATRSV 622

Qy    618 LHEVMEQQTVSVAKGGIIASLNARTSVLACANPSGSRYNARLSVIDNIQLPPTLLSRFDL 677
Db    623 LHEVMEQQTISIAKAGIITTLNARTSILASANPINSRYDPNLPVTGNIDLPPPLLSRFDL 682

Qy    678 IYLMLDKPDEQNDRRLARHLVALHYEN--YEVSKQDALDLQTLTAYITYARQHVHPTLSD 735
Db    683 VYLILDKVDESIDRQLARHLTDMYLEDAPETVNANSVLPVELLTLYIQYAKENFNPVMTE 742

Qy    736 EAAEDLINGYVEMRQKGNFPGSSKKVITATPRQLESMIRISEALARMRFSEVVEKVDAAE 795
Db    743 EGKNELVRSYVEMRKLGEDARSSEKRITATTRQLESMIRLSEAHAKMRLSERVELIDVKE 802

Qy    796 AVRLLDVALQQSATDHATGTIDMDLITTGVSASERIRRANLLAALRELIADKISPGSSSG 855
Db    803 AVRLIKSAIKDYATDPVTGRIDMDMIQTGTTAQQRRVQEDLVSEIMKIIEE-----NNNL 857

Qy    856 LKTSQLLEDIRSQSSVDVSLQDIKNALGSLQGEGFLTVHGDIVKR 900
Db    858 IRFNDLSVKLNERSSFRVENSVINEGLRRLQQEGKIMETGDSHRR 902
```

Search completed: December 7, 2005, 17:38:51
Job time : 199 secs

GenCore version 5.1.6


```
OM protein - protein search, using sw model

Run on:          December  7, 2005, 17:35:21 ; Search time 47 Seconds
                                              (without alignments)
                                              1584.910 Million cell updates/sec

Title:           US-10-768-511-8
Perfect score:   4609
Sequence:        1 MENNDALDIGAVSSPYPSQS..........LGSLQGEGFLTVHGDIVKRV 901

Scoring table:   BLOSUM62
                 Gapop 10.0 , Gapext 0.5

Searched:        572060 seqs, 82675679 residues

Total number of hits satisfying chosen parameters:      572060

Minimum DB seq length: 0
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%
                 Maximum Match 100%
                 Listing first 45 summaries

Database :       Issued_Patents_AA:*
                 1:  /cgn2_6/ptodata/1/iaa/5_COMB.pep:*
                 2:  /cgn2_6/ptodata/1/iaa/6_COMB.pep:*
                 3:  /cgn2_6/ptodata/1/iaa/H_COMB.pep:*
                 4:  /cgn2_6/ptodata/1/iaa/PCTUS_COMB.pep:*
                 5:  /cgn2_6/ptodata/1/iaa/RE_COMB.pep:*
                 6:  /cgn2_6/ptodata/1/iaa/backfiles1.pep:*
```

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	% Query Match	Length	DB	ID	Description
1	4609	100.0	901	2	US-09-828-062-8	Sequence 8, Appli
2	965.5	20.9	879	2	US-09-949-016-8492	Sequence 8492, Ap
3	963.5	20.9	720	2	US-09-604-605-2	Sequence 2, Appli
4	963.5	20.9	720	2	US-10-101-080-2	Sequence 2, Appli
5	946.5	20.5	904	2	US-09-964-899-49	Sequence 49, Appl
6	939	20.4	753	2	US-09-949-016-10892	Sequence 10892, A
7	936	20.3	909	2	US-09-976-594-324	Sequence 324, App
8	932	20.2	1017	2	US-09-538-092-319	Sequence 319, App
9	919.5	20.0	895	2	US-09-538-092-1158	Sequence 1158, Ap
10	916.5	19.9	823	2	US-09-949-016-8339	Sequence 8339, Ap
11	817.5	17.7	544	2	US-09-248-796A-18637	Sequence 18637, A
12	734.5	15.9	495	2	US-09-248-796A-20007	Sequence 20007, A
13	662	14.4	331	2	US-09-248-796A-18673	Sequence 18673, A
14	474	10.3	435	2	US-09-248-796A-18814	Sequence 18814, A
15	391.5	8.5	173	2	US-09-248-796A-20019	Sequence 20019, A
16	357	7.7	429	2	US-09-248-796A-19091	Sequence 19091, A
17	246.5	5.3	163	2	US-09-248-796A-20018	Sequence 20018, A
18	216	4.7	248	2	US-09-248-796A-19090	Sequence 19090, A
19	199	4.3	353	2	US-09-248-796A-20008	Sequence 20008, A
20	152.5	3.3	1781	2	US-09-961-403-13	Sequence 13, Appl
21	151	3.3	265	2	US-09-248-796A-18812	Sequence 18812, A
22	150	3.3	1493	2	US-09-423-890-8	Sequence 8, Appli
23	149.5	3.2	1492	2	US-09-697-898-5	Sequence 5, Appli
24	149	3.2	1780	1	US-08-769-309A-5	Sequence 5, Appli
25	149	3.2	1780	2	US-08-994-570-5	Sequence 5, Appli
26	142	3.1	105	2	US-09-513-999C-7279	Sequence 7279, Ap
27	142	3.1	1492	2	US-09-697-898-4	Sequence 4, Appli
28	141.5	3.1	1293	2	US-09-252-991A-17576	Sequence 17576, A
29	139.5	3.0	15281	1	US-08-471-119A-2	Sequence 2, Appli
30	138.5	3.0	679	2	US-09-248-796A-15353	Sequence 15353, A
31	138	3.0	1270	2	US-09-538-092-1321	Sequence 1321, Ap
32	137	3.0	2254	1	US-08-677-010-3	Sequence 3, Appli
33	137	3.0	2254	1	US-08-790-519-3	Sequence 3, Appli
34	136.5	3.0	832	2	US-09-540-236-3056	Sequence 3056, Ap
35	136.5	3.0	1593	2	US-08-628-829-4	Sequence 4, Appli
36	131.5	2.9	772	2	US-09-107-532A-5724	Sequence 5724, Ap
37	131	2.8	499	2	US-09-328-352-5415	Sequence 5415, Ap
38	129.5	2.8	852	2	US-09-741-150-2	Sequence 2, Appli
39	129.5	2.8	852	2	US-10-160-187-2	Sequence 2, Appli
40	127	2.8	943	2	US-09-902-540-10641	Sequence 10641, A
41	127	2.8	2032	2	US-09-071-035-458	Sequence 458, App
42	127	2.8	2032	2	US-09-071-035-462	Sequence 462, App
43	127	2.8	2032	2	US-09-071-035-466	Sequence 466, App
44	127	2.8	2032	2	US-10-206-576-458	Sequence 458, App
45	127	2.8	2032	2	US-10-206-576-462	Sequence 462, App

ALIGNMENTS

```
RESULT 1
US-09-828-062-8
; Sequence 8, Application US/09828062
; Patent No. 6710229
; GENERAL INFORMATION:
;  APPLICANT: COSTA E SILVA, OSWALDO DA
;  APPLICANT:  BOHNERT, HANS J.
;  APPLICANT:  VAN THIELEN, NOCHA
;  APPLICANT:  CHEN, ROUYING
;  APPLICANT:  SARRIA-MILLAN, RODRIGO
;  TITLE OF INVENTION: CELL CYCLE STRESS-RELATED PROTEINS AND METHODS OF USE
;  TITLE OF INVENTION:  IN PLANTS
;  FILE REFERENCE: 16313-0031
;  CURRENT APPLICATION NUMBER: US/09/828,062
;  CURRENT FILING DATE:  2001-08-20
;  PRIOR APPLICATION NUMBER: 60/196,001
;  PRIOR FILING DATE: 2000-04-07
;  NUMBER OF SEQ ID NOS: 34
;  SOFTWARE: PatentIn Ver. 2.1
; SEQ ID NO 8
;   LENGTH: 901
;   TYPE: PRT
;   ORGANISM: Physcomitrella patens
US-09-828-062-8

  Query Match             100.0%;  Score 4609;  DB 2;  Length 901;
  Best Local Similarity   100.0%;  Pred. No. 0;
  Matches  901;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy          1 MENNDALDIGAVSSPYPSQSEGVSTPLPQVTSPSFDNAASPVAGRRAVRQTPTSAVRRRG 60
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db          1 MENNDALDIGAVSSPYPSQSEGVSTPLPQVTSPSFDNAASPVAGRRAVRQTPTSAVRRRG 60

Qy         61 RETDSARRRRSRSRSLGNSVYSSPYDAGTPGTPGTPVATPVYATPVGTPMGTPSFHRGTP 120
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db         61 RETDSARRRRSRSRSLGNSVYSSPYDAGTPGTPGTPVATPVYATPVGTPMGTPSFHRGTP 120

Qy        121 QYKQRSELGSQGKPLHRRRRSQSREPGHRSPSREPSADGRPSESAEPDDTLGGEYAYVWG 180
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        121 QYKQRSELGSQGKPLHRRRRSQSREPGHRSPSREPSADGRPSESAEPDDTLGGEYAYVWG 180

Qy        181 TNVNIPDVLRAIRRFLHNYRSSAHDLNSKYIQIIEETVEREEDTLNIDMSDIYDHDPDLY 240
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        181 TNVNIPDVLRAIRRFLHNYRSSAHDLNSKYIQIIEETVEREEDTLNIDMSDIYDHDPDLY 240

Qy        241 AKIVRYPLDIIPLLDTECQEVATSLLPTFEKHIEARPFNLKASVHMRELNPSDIDKLVSV 300
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        241 AKIVRYPLDIIPLLDTECQEVATSLLPTFEKHIEARPFNLKASVHMRELNPSDIDKLVSV 300

Qy        301 KGMVIRCSSIIPEIKGAFFKCLVCGHSPPLVTVVKGRVEEPTRCEKPECAARNAMSLIHN 360
```

```
                 ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db          301 KGMVIRCSSIIPEIKGAFFKCLVCGHSPPLVTVVKGRVEEPTRCEKPECAARNAMSLIHN 360

Qy          361 RCTFANKQIVRLQETPDAIPEGETPHTVSMCLYNTMVDAVKPGDRIEVTGVFKAMAVRVG 420
                 ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db          361 RCTFANKQIVRLQETPDAIPEGETPHTVSMCLYNTMVDAVKPGDRIEVTGVFKAMAVRVG 420

Qy          421 PNQRTLRALYKTYIDCVHVKKSDRGRLQTEDPMEMDKENDMYAGYHESDTSEAANEAKIQ 480
                 ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db          421 PNQRTLRALYKTYIDCVHVKKSDRGRLQTEDPMEMDKENDMYAGYHESDTSEAANEAKIQ 480

Qy          481 KLKELSKLPGIYDRLSRSLAPSIWELEDIKKGLLCQLFGGKAKKIPSGASFRGDINVLLV 540
                 ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db          481 KLKELSKLPGIYDRLSRSLAPSIWELEDIKKGLLCQLFGGKAKKIPSGASFRGDINVLLV 540

Qy          541 GDPGTSKSQLLQYVHKIAPRGIYTSGRGSSAVGLTAYVTKDPETRETVLESGALVLSDRG 600
                 ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db          541 GDPGTSKSQLLQYVHKIAPRGIYTSGRGSSAVGLTAYVTKDPETRETVLESGALVLSDRG 600

Qy          601 ICCIDEFDKMSDNARSMLHEVMEQQTVSVAKGGIIASLNARTSVLACANPSGSRYNARLS 660
                 ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db          601 ICCIDEFDKMSDNARSMLHEVMEQQTVSVAKGGIIASLNARTSVLACANPSGSRYNARLS 660

Qy          661 VIDNIQLPPTLLSRFDLIYLMLDKPDEQNDRRLARHLVALHYENYEVSKQDALDLQTLTA 720
                 ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db          661 VIDNIQLPPTLLSRFDLIYLMLDKPDEQNDRRLARHLVALHYENYEVSKQDALDLQTLTA 720

Qy          721 YITYARQHVHPTLSDEAAEDLINGYVEMRQKGNFPGSSKKVITATPRQLESMIRISEALA 780
                 ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db          721 YITYARQHVHPTLSDEAAEDLINGYVEMRQKGNFPGSSKKVITATPRQLESMIRISEALA 780

Qy          781 RMRFSEVVEKVDAAEAVRLLDVALQQSATDHATGTIDMDLITTGVSASERIRRANLLAAL 840
                 ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db          781 RMRFSEVVEKVDAAEAVRLLDVALQQSATDHATGTIDMDLITTGVSASERIRRANLLAAL 840

Qy          841 RELIADKISPGSSSGLKTSQLLEDIRSQSSVDVSLQDIKNALGSLQGEGFLTVHGDIVKR 900
                 ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db          841 RELIADKISPGSSSGLKTSQLLEDIRSQSSVDVSLQDIKNALGSLQGEGFLTVHGDIVKR 900

Qy          901 V 901
                 |
Db          901 V 901
```

RESULT 2
US-09-949-016-8492
; Sequence 8492, Application US/09949016
; Patent No. 6812339
; GENERAL INFORMATION:
; APPLICANT: VENTER, J. Craig et al.
; TITLE OF INVENTION: POLYMORPHISMS IN KNOWN GENES ASSOCIATED
; TITLE OF INVENTION: WITH HUMAN DISEASE, METHODS OF DETECTION AND USES THEREOF
; FILE REFERENCE: CL001307
; CURRENT APPLICATION NUMBER: US/09/949,016
; CURRENT FILING DATE: 2000-04-14
; PRIOR APPLICATION NUMBER: 60/241,755
; PRIOR FILING DATE: 2000-10-20
; PRIOR APPLICATION NUMBER: 60/237,768
; PRIOR FILING DATE: 2000-10-03
; PRIOR APPLICATION NUMBER: 60/231,498
; PRIOR FILING DATE: 2000-09-08
; NUMBER OF SEQ ID NOS: 207012
; SOFTWARE: FastSEQ for Windows Version 4.0
; SEQ ID NO 8492
; LENGTH: 879
; TYPE: PRT
; ORGANISM: Human
US-09-949-016-8492

```
  Query Match             20.9%;  Score 965.5;  DB 2;  Length 879;
  Best Local Similarity   30.9%;  Pred. No. 5.1e-76;
  Matches 264;  Conservative 143;  Mismatches 249;  Indels 197;  Gaps 30;

Qy          130 SQGKPLHRRRRSQSREPGHRSPSREPSADGRPSESAEPDDTLGGEYAYVWGTNVNIPD-V 188
Db           38 SVGRPQSTARRF-SASPGAR-PLSPPSAAARPSAAMALKD-----YALEKGTGLRAREGV 90

Qy          189 LRA-----------------------------------IRRFLHNYRSSAHDLNSKY 210
Db           91 LRGEPPGEHLLFIGRVLSENGYLRFVLKFVSYGFTFHVTEKVKKFLQEFYQD-DELGKKQ 149

Qy          211 IQIIEETV---EREEDTLNIDMSDIYDHDPDL----------YAKIVRYPL-DIIP---- 252
Db          150 FKYGNQLVRLAHREQVALYVDLDDVAEDDPELVDSICENARRYAKLFADAVQELLPQYKE 209

Qy          253 -------LLDTECQ-------------------EVATSLLPTFEKHIEARPFNLKASVH 285
Db          210 REVVNKDVLDVYIEHRLMMEQRSRDPGMVRSPQNQYPAELMRRFELYFQG-PSSNKPRV- 267

Qy          286 MRELNPSDIDKLVSVKGMVIRCSSIIPEIKGAFFKCLVCG---HSPPLVTVVKGRVEEPT 342
Db          268 IREVRADSVGKLVTVRGIVTRVSEVKPKMVVATYTCDQCGAETYQP---------IQSPT 318

Qy          343 -----RCEKPECAARNA---MSLIHNRCTFANKQIVRLQETPDAIPEGETPHTVSMCLYN 394
Db          319 FMPLIMCPSQECQTNRSGGRLYLQTRGSRFIKFQEMKMQEHSDQVPVGNIPRSITVLVEG 378

Qy          395 TMVDAVKPGDRIEVTGVFKAMAVRVG------------PNQ-RT---------------L 426
Db          379 ENTRIAQPGDHVSVTGIFLPI-LRTGFRQVVQRKEQQNPKQNRTQGFVKKLLGNPVMGKL 437

Qy          427 RALY------------KTYIDCVHVKKSDRGRLQTEDPMEMDKENDMYAGYHESDTSEA 473
Db          438 KGLHLYPGIFLPKGLLSETYLEAHRIVK------------MNKSEDDESG--------- 475

Qy          474 ANEAKIQKLKELSKLPGIYDRLSRSLAPSIWELEDIKKGLLCQLFGGKAKKIPSGASFRG 533
Db          476 AGELTREELRQIAE-EDFYEKLAASIAPEIYGHEDVKKALLLLLVGG-VDQSPRGMKIRG 533

Qy          534 DINVLLVGDPGTSKSQLLQYVHKIAPRGIYTSGRGSSAVGLTAYVTKDPETRETVLESGA 593
Db          534 NINICLMGDPGVAKSQLLSYIDRLAPRSQYTTGRGSSGVGLTAAVLRDSVSGELTLEGGA 593

Qy          594 LVLSDRGICCIDEFDKMSDNARSMLHEVMEQQTVSVAKGGIIASLNARTSVLACANPSGS 653
Db          594 LVLADQGVCCIDEFDKMAEADRTAIHEVMEQQTISIAKAGILTTLNARCSILAAANPAYG 653

Qy          654 RYNARLSVIDNIQLPPTLLSRFDLIYLMLDKPDEQNDRRLARHLVALHYENYEVSKQ-DA 712
Db          654 RYNPRRSLEQNIQLPAALLSRFDLLWLIQDRPDRDNDLRLAQHITYVHQHSRQPPSQFEP 713

Qy          713 LDLQTLTAYITYARQHVHPTLSDEAAEDLINGYVEMRQKGNFPGSSKKVITATPRQLESM 772
Db          714 LDMKLMRRYIAMCREK-QPMVPESLADYITAAYVEMRREA---WASKDATYTSARTLLAI 769

Qy          773 IRISEALARMRFSEVVEKVDAAEAVRLLDVALQQSATDHATGTIDMDLITTGVSASERIR 832
Db          770 LRLSTALARLRMVDVVEKEDVNEAIRLMEMSKDSLLGD-----------KGQTARTQRP 817

Qy          833 RANLLAALRELIA 845
Db          818 ADVIFATVRELVS 830
```

RESULT 3
US-09-604-605-2
; Sequence 2, Application US/09604605
; Patent No. 6421613
; GENERAL INFORMATION:
; APPLICANT: Ramgopal Nadimpalli
; APPLICANT: Carl R. Simmons
; TITLE OF INVENTION: Maize Prolifera Gene and Uses Thereof
; FILE REFERENCE: 1138
; CURRENT APPLICATION NUMBER: US/09/604,605
; CURRENT FILING DATE: 2000-06-27
; PRIOR APPLICATION NUMBER: 60/143,222

```
; PRIOR FILING DATE: 1999-07-09
; NUMBER OF SEQ ID NOS: 2
; SOFTWARE: FastSEQ for Windows Version 3.0
; SEQ ID NO 2
;   LENGTH: 720
;   TYPE: PRT
;   ORGANISM: Zea mays
US-09-604-605-2

  Query Match             20.9%;  Score 963.5;  DB 2;  Length 720;
  Best Local Similarity   34.3%;  Pred. No. 5.5e-76;
  Matches 256;  Conservative 125;  Mismatches 241;  Indels 125;  Gaps 22;

Qy    195 FLHNYRSSAHDLNSKYIQIIEETVEREEDTLNIDMSDIY---DHDPDLYAKIV----RY- 246
Db     16 FLSNF-TGPHG-EPKYLNILQDVANRKIRAVQIELDDLFHYKDVDEEFLQRVTENTRRYI 73

Qy    247 -----PLDIIPLLDTEC----------------------------QEVATSLLPTFEKH 272
Db     74 GIFAEAMDEIMPEPTEAYTVDEDQDILMTQRVDEGADGGADGTDPLQKMPPEIKRFFEVY 133

Qy    273 IEARPFNLKASVHMRELNPSDIDKLVSVKGMVIRCSSIIPEIKGAFFKCLVCGHSPPLVT 332
Db    134 IKA--FSKVTPLTIRQVKASNIGQLVKISGIVTRCSDVKPLMQVAVYTCEECGFE--IYQ 189

Qy    333 VVKGRVEEP-TRCEKPEC---AARNAMSLIHNRCTFANKQIVRLQETPDAIPEGETPHTV 388
Db    190 EVTARVFMPLIECPSQRCKLNKAKGNLILQLRASKFLKFQEVKLQELAEHVPKGHIPRAL 249

Qy    389 SMCLYNTMVDAVKPGDRIEVTGVFKAM------AVRVGPNQRTLRALYKTYIDCV---HV 439
Db    250 TVHLRGELTRKVAPGDVVEMSGIFLPMPYYGFRAMRAG-------LVADTYLEAMSVTHF 302

Qy    440 KKSDRGRLQTEDPMEMDKENDMYAGYHESDTSEAANEAKIQKLKELSKLPGIYDRLSRSL 499
Db    303 KKK---------------------YEEYDLKGDEQE----QIDRLAEDGDIYSKLARSL 336

Qy    500 APSIWELEDIKKGLLCQLFGGKAKKIPSGASFRGDINVLLVGDPGTSKSQLLQYVHKIAP 559
Db    337 APEIFGHEDVKKALLLLLVGAPHRKLADGMKIRGDLHICMMGDPGVAKSQLLKHIINVAP 396

Qy    560 RGIYTSGRGSSAVGLTAYVTKDPETRETVLESGALVLSDRGICCIDEFDKMSDNARSMLH 619
Db    397 RGVYTTGRGSSGVGLTAAVQKDPVTNEFVLEGGALVLADMGICAIDEFDKMEESDRTAIH 456

Qy    620 EVMEQQTVSVAKGGIIASLNARTSVLACANPSGSRYNARLSVIDNIQLPPTLLSRFDLIY 679
Db    457 EVMEQQTVSIAKAGITTSLNARTAILAAANPAWGRYDMRRTPAENINLPPALLSRFDLLW 516

Qy    680 LMLDKPDEQNDRRLARHLVALHYENYEVSKQ--DALDLQTLTAYITYARQHVHPTLSDEA 737
Db    517 LILDRADMETDLEMARHVVHVH-QNLESPALGFTPLEPSVLRAYISAARR-VIPSVPREL 574

Qy    738 AEDLINGYVEMRQ---KGNFPGSSKKVITATPRQLESMIRISEALARMRFSEVVEKVDAA 794
Db    575 EEYIATAYSSIRQEEAKSNAPTS-----YTTIRTLLSILRISIALARLRFSETVAQSDVD 629

Qy    795 EAVRLLDVA--------LQQSATDHATGTIDMDLITTGVSASERIRRANLLAALRELIAD 846
Db    630 EALRLMQMSKYSLYSDDRQRSGLDAISDIYSILRDEAARTSSMDVRYAHALNLI------ 683

Qy    847 KISPGSSSGLKTSQLLEDIRSQSSVDV 873
Db    684 -----SRKGYSEAQLKECLEEYASLNV 705

RESULT 4
US-10-101-080-2
; Sequence 2, Application US/10101080
; Patent No. 6879916
; GENERAL INFORMATION:
;  APPLICANT: Ramgopal Nadimpalli
;  APPLICANT: Carl R. Simmons
; TITLE OF INVENTION: Maize Prolifera Gene and Uses Thereof
; FILE REFERENCE: 1138
; CURRENT APPLICATION NUMBER: US/10/101,080
; CURRENT FILING DATE: 2002-05-09
; PRIOR APPLICATION NUMBER: US/09/604,605
; PRIOR FILING DATE: 2000-06-27
; NUMBER OF SEQ ID NOS: 2
; SOFTWARE: FastSEQ for Windows Version 3.0
; SEQ ID NO 2
;   LENGTH: 720
;   TYPE: PRT
;   ORGANISM: Zea mays
US-10-101-080-2

  Query Match             20.9%;  Score 963.5;  DB 2;  Length 720;
  Best Local Similarity   34.3%;  Pred. No. 5.5e-76;
  Matches 256;  Conservative 125;  Mismatches 241;  Indels 125;  Gaps 22;

Qy    195 FLHNYRSSAHDLNSKYIQIIEETVEREEDTLNIDMSDIY---DHDPDLYAKIV----RY- 246
Db     16 FLSNF-TGPHG-EPKYLNILQDVANRKIRAVQIELDDLFHYKDVDEEFLQRVTENTRRYI 73

Qy    247 -----PLDIIPLLDTEC----------------------------QEVATSLLPTFEKH 272
Db     74 GIFAEAMDEIMPEPTEAYTVDEDQDILMTQRVDEGADGGADGTDPLQKMPPEIKRFFEVY 133

Qy    273 IEARPFNLKASVHMRELNPSDIDKLVSVKGMVIRCSSIIPEIKGAFFKCLVCGHSPPLVT 332
Db    134 IKA--FSKVTPLTIRQVKASNIGQLVKISGIVTRCSDVKPLMQVAVYTCEECGFE--IYQ 189

Qy    333 VVKGRVEEP-TRCEKPEC---AARNAMSLIHNRCTFANKQIVRLQETPDAIPEGETPHTV 388
Db    190 EVTARVFMPLIECPSQRCKLNKAKGNLILQLRASKFLKFQEVKLQELAEHVPKGHIPRAL 249

Qy    389 SMCLYNTMVDAVKPGDRIEVTGVFKAM------AVRVGPNQRTLRALYKTYIDCV---HV 439
Db    250 TVHLRGELTRKVAPGDVVEMSGIFLPMPYYGFRAMRAG-------LVADTYLEAMSVTHF 302

Qy    440 KKSDRGRLQTEDPMEMDKENDMYAGYHESDTSEAANEAKIQKLKELSKLPGIYDRLSRSL 499
Db    303 KKK---------------------YEEYDLKGDEQE----QIDRLAEDGDIYSKLARSL 336

Qy    500 APSIWELEDIKKGLLCQLFGGKAKKIPSGASFRGDINVLLVGDPGTSKSQLLQYVHKIAP 559
Db    337 APEIFGHEDVKKALLLLLVGAPHRKLADGMKIRGDLHICMMGDPGVAKSQLLKHIINVAP 396

Qy    560 RGIYTSGRGSSAVGLTAYVTKDPETRETVLESGALVLSDRGICCIDEFDKMSDNARSMLH 619
Db    397 RGVYTTGRGSSGVGLTAAVQKDPVTNEFVLEGGALVLADMGICAIDEFDKMEESDRTAIH 456

Qy    620 EVMEQQTVSVAKGGIIASLNARTSVLACANPSGSRYNARLSVIDNIQLPPTLLSRFDLIY 679
Db    457 EVMEQQTVSIAKAGITTSLNARTAILAAANPAWGRYDMRRTPAENINLPPALLSRFDLLW 516

Qy    680 LMLDKPDEQNDRRLARHLVALHYENYEVSKQ--DALDLQTLTAYITYARQHVHPTLSDEA 737
Db    517 LILDRADMETDLEMARHVVHVH-QNLESPALGFTPLEPSVLRAYISAARR-VIPSVPREL 574

Qy    738 AEDLINGYVEMRQ---KGNFPGSSKKVITATPRQLESMIRISEALARMRFSEVVEKVDAA 794
Db    575 EEYIATAYSSIRQEEAKSNAPTS-----YTTIRTLLSILRISIALARLRFSETVAQSDVD 629

Qy    795 EAVRLLDVA--------LQQSATDHATGTIDMDLITTGVSASERIRRANLLAALRELIAD 846
Db    630 EALRLMQMSKYSLYSDDRQRSGLDAISDIYSILRDEAARTSSMDVRYAHALNLI------ 683

Qy    847 KISPGSSSGLKTSQLLEDIRSQSSVDV 873
Db    684 -----SRKGYSEAQLKECLEEYASLNV 705

RESULT 5
US-09-964-899-49
```

```
; Sequence 49, Application US/09964899
; Patent No. 6900367
; GENERAL INFORMATION:
;  APPLICANT: Cohen, Dalia et al.
;  TITLE OF INVENTION: Identification of Genes Involved in
;  TITLE OF INVENTION:  Alzheimer's Disease Using Drosophila Melanogaster
;  FILE REFERENCE: 4-31612 A
;  CURRENT APPLICATION NUMBER: US/09/964,899
;  CURRENT FILING DATE:  2001-09-27
;  PRIOR APPLICATION NUMBER: 60/236,893
;  PRIOR FILING DATE: 2000-09-29
;  PRIOR APPLICATION NUMBER: 60/298,309
;  PRIOR FILING DATE: 2001-06-14
;  NUMBER OF SEQ ID NOS: 53
;  SOFTWARE: FastSEQ for Windows Version 4.0
; SEQ ID NO 49
;   LENGTH: 904
;   TYPE: PRT
;   ORGANISM: Homo Sapien
US-09-964-899-49

  Query Match             20.5%;  Score 946.5;  DB 2;  Length 904;
  Best Local Similarity   29.0%;  Pred. No. 2.6e-74;
  Matches  262;  Conservative  157;  Mismatches  344;  Indels  139;  Gaps   23;

Qy         29 QVTSPSFDNAASPVAGRRAVRQTPTSAVRRRGRETDSARRRRSRSRSLGNSVYSSPYDAG 88
Db          2 QESSESFTMASSP-AQRRRGNDPLTSSPGRSSRRTDAL--TSSPGRDL------PPFEDE 52

Qy         89 TPGTPGTPVATPVYATPVGTPM---GTPSFHRGTPQYKQRSELG-------------SQG 132
Db         53 SEGLLGT--EGPLEEEEDGEELIGDGMERDYRAIPELDAYEAEGLALDDEDVEELTASQR 110

Qy        133 KPLHRRRRSQSREPG-----------------HRSPSREPSADGRPSESAEPD------ 168
Db        111 EAAERAMRQRDREAGRGLGRMRRGLLYDSDEEDEERPARKRRQVERATEDGEEDEEMIES 170

Qy        169 ----DTLGGEYAYVW----GTNVNIPDVLRAIRRFLHNY-RSSAHDLNSKYIQIIEETVE 219
Db        171 IENLEDLKGHSVREWVSMAGPRL---EIHHRFKNFLRTHVDSHGHNV---FKERISDMCK 224

Qy        220 REEDTLNIDMSDIYDHDPDLYAKIVRYPLDIIPLLDTECQEVATSLLPTFEK---HIEAR 276
Db        225 ENRESLVVNYEDLAAREHVLAYFLPEAPAELLQIFDEAALEVVLAMYPKYDRITNHIHVR 284

Qy        277 PFNLKASVHMRELNPSDIDKLVSVKGMVIRCSSIIPEIKGAFFKCLVCGHSPPLVTVVKG 336
Db        285 ISHLPLVEELRSLRQLHLNQLIRTSGVVTSCTGVLPQLSMVKYNCNKCNFVLGPFCQSQN 344

Qy        337 RVEEPTRCEKPECAARNAMSLIHNRCTFANKQIVRLQETPDAIPEGETPHTVSMCLYNTM 396
Db        345 QEVKPGSC--PECQSAGPFEVNMEETIYQNYQRIRIQESPGKVAAGRLPRSKDAILLADL 402

Qy        397 VDAVKPGDRIEVTGVFKAMAVRVGPNQRTLRA--LYKTYIDCVHVKKSDR----GRLQTE 450
Db        403 VDSCKPGDEIELTGIYHN---NYDGSLNTANGFPVFATVILANHVAKKDNKVAVGELTDE 459

Qy        451 DPMEMDKENDMYAGYHESDTSEAANEAKIQKLKELSKLPGIYDRLSRSLAPSIWELEDIK 510
Db        460 D--------------------------VKMITSLSKDQQIGEKIFASIAPSIYGHEDIK 492

Qy        511 KGLLCQLFGGKAKKIPSGASFRGDINVLLVGDPGTSKSQLLQYVHKIAPRGIYTSGRGSS 570
Db        493 RGLALALFGGEPKNPGGKHKVRGDINVLLCGDPGTAKSQFLKYIEKVSSRAIFTTGQGAS 552

Qy        571 AVGLTAYVTKDPETRETVLESGALVLSDRGICCIDEFDKMSDNARSMLHEVMEQQTVSVA 630
Db        553 AVGLTAYVQRHPVSREWTLEAGALVLADRGVCLIDEFDKMNDQDRTSIHEAMEQQSISIS 612

Qy        631 KGGIIASLNARTSVLACANPSGSRYNARLSVIDNIQLPPTLLSRFDLIYLMLDKPDEQND 690
Db        613 KAGIVTSLQARCTVIAAANPIGGRYDPSLTFSENVDLTEPIISRFDILCVVRDTVDPVQD 672

Qy        691 RRLARHLVALHYENYEVSKQ-------------------DALDLQTLTAYITYARQHVH 730
Db        673 EMLARFVVGSHVRHHPSNKEEEGLANGSAAEPAMPNTYGVEPLPQEVLKKYIIYAKERVH 732

Qy        731 PTLSDEAAEDLINGYVEMRQKGNFPGSSKKVITATPRQLESMIRISEALARMRFSEVVEK 790
Db        733 PKLNQMDQDKVAKMYSDLRKESMATGS----IPITVRHIESMIRMAEAHARIHLRDYVIE 788

Qy        791 VDAAEAVRLLDVALQQSATDHATGTIDMDLITTGVSASERIRRAN---LLAALRELIADK 847
Db        789 DDVNMAIRV----MLESFIDTQKFSVMRSMRKT-FARYLSFRRDNNELLLFILKQLVAEQ 843

Qy        848 IS 849
Db        844 VT 845

RESULT 6
US-09-949-016-10892
; Sequence 10892, Application US/09949016
; Patent No. 6812339
; GENERAL INFORMATION:
;  APPLICANT: VENTER, J. Craig et al.
;  TITLE OF INVENTION: POLYMORPHISMS IN KNOWN GENES ASSOCIATED
;  TITLE OF INVENTION:  WITH HUMAN DISEASE, METHODS OF DETECTION AND USES THEREOF
;  FILE REFERENCE: CL001307
;  CURRENT APPLICATION NUMBER: US/09/949,016
;  CURRENT FILING DATE:  2000-04-14
;  PRIOR APPLICATION NUMBER: 60/241,755
;  PRIOR FILING DATE: 2000-10-20
;  PRIOR APPLICATION NUMBER: 60/237,768
;  PRIOR FILING DATE: 2000-10-03
;  PRIOR APPLICATION NUMBER: 60/231,498
;  PRIOR FILING DATE: 2000-09-08
;  NUMBER OF SEQ ID NOS: 207012
;  SOFTWARE: FastSEQ for Windows Version 4.0
; SEQ ID NO 10892
;   LENGTH: 753
;   TYPE: PRT
;   ORGANISM: Human
US-09-949-016-10892

  Query Match             20.4%;  Score 939;  DB 2;  Length 753;
  Best Local Similarity   30.9%;  Pred. No. 8.7e-74;
  Matches  232;  Conservative  156;  Mismatches  290;  Indels   72;  Gaps   18;

Qy        169 DTLGGEYAYVWGTNVNIPDVLRAIRRFLHNYR--SSAHDLNSKYIQIIEETVEREEDTLN 226
Db         32 DSFGGD-AQADEGQARKSQLQRRFKEFLRQYRVGTDRTGFTFKYRDELKRHYNLGEYWIE 90

Qy        227 IDMSDIYDHDPDLYAKIVRYPLDIIPLLDTECQEVA---TSLLPTFEKHIEARPFNLKAS 283
Db         91 VEMEDLASFDEDLADYLYKQPAEHLQLLEEAAKEVADEVTRPRPSGEEVLQDIQVMLKSD 150

Qy        284 V---HMRELNPSDIDKLVSVKGMVIRCSSIIPEIKGAFFKCLVCGHSPPLVTVVKG--RV 338
Db        151 ASPSSIRSLKSDMMSHLVKIPGIIIAASAVRAKATRISIQCRSCRNTLTNIAMRPGLEGY 210

Qy        339 EEPTRCE-----KPECAARNAMSLIHNRCTFANKQIVRLQETPDAIPEGETPHTVSMCLY 393
Db        211 ALPRKCNTDQAGRPKCPL-DPYFIMPDKCKCVDFQTLKLQELPDAVPHGEMPRHMQLYCD 269

Qy        394 NTMVDAVKPGDRIEVTGVFKAMAV---------RVGPNQRTLRALYKTYIDCVHVKKSDR 444
Db        270 RYLCDKVVPGNRVTIMGIYSIKKFGLTTSRGRDRVGVGIRS------SYIRVLGIQ---- 319

Qy        445 GRLQTEDPMEMDKENDMYAGYHESDTSEAANEAKIQKLKELSKLPGIYDRLSRSLAPSIW 504
Db        320 --------VDTDGSGRSFAG--------AVSPQEEEEFRRLAALPNVYEVISKSIAPSIF 363

Qy        505 ELEDIKKGLLCQLFGGKAKKIPSGASFRGDINVLLVGDPGTSKSQLLQYVHKIAPRGIYT 564
Db        364 GGTDMKKAIACLLFGGSRKRLPDGLTRRGDINLLMLGDPGTAKSQLLKFVEKCSPIGVYT 423
```

```
Qy        565 SGRGSSAVGLTAYVTKDPETRETVLESGALVLSDRGICCIDEFDKMSDNARSMLHEVMEQ 624
              ||:|||| |||| | :|| :|   ::| ||:||:| |: ||||||| :: |  :|| |||
Db        424 SGKGSSAAGLTASVMRDPSSRNFIMEGGAMVLADGGVVCIDEFDKMREDDRVAIHEAMEQ 483

Qy        625 QTVSVAKGGIIASLNARTSVLACANPSGSRYNARLSVIDNIQLPPTLLSRFDLIYLMLDK 684
              ||:|:|| ||  :||:| |||| ||       |::        |||     ||:||||||:|:: |:
Db        484 QTISIAKAGITTTLNSRCSVLAAANSVFGRWDETKGE-DNIDFMPTILSRFDMIFIVKDE 542

Qy        685 PDEQNDRRLARHLVALHYENYEVSK--QDALDLQTLTAYITYARQHVHPTLSDEAAEDLI 742
               :|: |  ||:|:: ||       ::  :  :||  |  :| | |      || |||| |
Db        543 HNEERDVMLAKHVITLHVSALTQTQAVEGEIDLAKLKKFIAYCRVKCGPRLSAEAAEKLK 602

Qy        743 NGYVEMR---QKGNFPGSSKKVITATPRQLESMIRISEALARMRFSEVVEKVDAAEAVRL 799
              | |: ||    ::           :  |  | ||||:::||:|||::|:        :  |  ||:||
Db        603 NRYIIMRSGARQHERDSDRRSSIPITVRQLEAIVRIAEALSKMKLQPFATEADVEEALRL 662

Qy        800 LDVALQQSATDHATGTIDMDLITTGVSASERIRRANLLAALRELIADKISPGSSSGLKTS 859
               |:  :|    :||:       :||        :|: : : :  : :  ||   :
Db        663 FQVSTLDAA---LSGTL------SGVEGFTSQEDQEMLSRIEKQLKRRFAIGSQ--VSEH 711

Qy        860 QLLEDIRSQSSVDVSLQDIKNAL---GSLQ 886
               :::|   |   : ::  :  :    | :|
Db        712 SIIKDFTKQKYPEHAIHKVLQLMLRRGEIQ 741
```

RESULT 7
US-09-976-594-324
; Sequence 324, Application US/09976594
; Patent No. 6673549
; GENERAL INFORMATION:
; APPLICANT: Furness, Michael
; APPLICANT: Buchbinder, Jenny
; TITLE OF INVENTION: GENES EXPRESSED IN C3A LIVER CELL CULTURES TREATED WITH STEROIDS
; FILE REFERENCE: PA-0041 US
; CURRENT APPLICATION NUMBER: US/09/976,594
; CURRENT FILING DATE: 2001-10-12
; PRIOR APPLICATION NUMBER: 60/240,409
; PRIOR FILING DATE: 2000-10-12
; NUMBER OF SEQ ID NOS: 1143
; SOFTWARE: PERL Program
; SEQ ID NO 324
;   LENGTH: 909
;   TYPE: PRT
;   ORGANISM: Homo sapiens
;   FEATURE:
;   NAME/KEY: misc_feature
;   OTHER INFORMATION: Incyte ID No. 6673549 1723834CD1
US-09-976-594-324

```
  Query Match             20.3%;  Score 936;  DB 2;  Length 909;
  Best Local Similarity   28.8%;  Pred. No. 2.2e-73;
  Matches  261;  Conservative  159;  Mismatches  341;  Indels  144;  Gaps   24;

Qy         31 TSPSFDNAASPVAGRRAVRQTPTSAVRRRGRETDSARRRRSRSRSLGNSVYSSPYDAGTP 90
              :| ||  |:|| | ||        ||:  |  | ||:      |  | |          |:: :
Db          4 SSESFTMASSP-AQRRRGNDPLTSSPGRSSRRTDAL--TSSPGRDL------PPFEDESE 54

Qy         91 GTPGTPVATPVYATPVGTPM---GTPSFHRGTPQYKQRSELG------------SQGKP 134
              | ||  |:   |  :   |  :| |:           |            || :
Db         55 GLLGT--EGPLEEEEDGEELIGDGMERDYRAIPELDAYEAEGLALDDEDVEELTASQREA 112

Qy        135 LHRRRRSQSREPG-----------------HRSPSREPSADGRPSESAEPD-------- 168
               |  | : |||                  |:|:     |  :|  |
Db        113 AERAMRQRDREAGRGLGRMRRGLLYDSDEEDEERPARKRRQVERATEDGEEDEEMIESIE 172

Qy        169 --DTLGGEYAYVW----GTNVNIPDVLRAIRRFLHNY-RSSAHDLNSKYIQIIEETVERE 221
                : | |         |  :  ::   : ||  :  |  |::   : : | :  :
Db        173 NLEDLKGHSVREWVSMAGPRL---EIHHRFKNFLRTHVDSHGHNV---FKERISDMCKEN 226

Qy        222 EDTLNIDMSDIYDHDPDLYAKIVRYPLDIIPLLDTECQEVATSLLPTFEK---HIEARPF 278
               ::| ::  |:   :   |    :   | ::: : |     ||  :: | :::   || |
Db        227 RESLVVNYEDLAAREHVLAYFLPEAPAELLQIFDEAALEVVLAMYPKYDRITNHIHVRIS 286

Qy        279 NLKASVHMRELNPSD-----IDKLVSVKGMVIRCSSIIPEIKGAFFKCLVCGHSPPLVTV 333
              :|     :| | ::      :::|:   |:|  |: ::|::   : |  |
Db        287 HLPLVEELRSLRRAEGRQLHLNQLIRTSGVVTSCTGVLPQLSMVKYNCNKCNFVLGPFCQ 346

Qy        334 VKGRVEEPTRCEKPECAARNAMSLIHNRCTFANKQIVRLQETPDAIPEGETPHTVSMCLY 393
               : :  :|  |  ||| :      :          : | | :|:||:|   :  |  | :   |
Db        347 SQNQEVKPGSC--PECQSAGPFEVNMEETIYQNYQRIRIQESPGKVAAGRLPRSKDAILL 404

Qy        394 NTMVDAVKPGDRIEVTGVFKAMAVRVGPNQRTLRA--LYKTYIDCVHVKKSDR----GRL 447
               :||: |||| ||:||::              :  |      :: | |   || | |      | |
Db        405 ADLVDSCKPGDEIELTGIYHN---NYDGSLNTANGFPVFATVILANHVAKKDNKVAVGEL 461

Qy        448 QTEDPMEMDKENDMYAGYHESDTSEAANEAKIQKLKELSKLPGIYDRLSRSLAPSIWELE 507
              ||                           :: :  |||   | :::  |:||||:  |
Db        462 TDED--------------------------VKMITSLSKDQQIGEKIFASIAPSIYGHE 494

Qy        508 DIKKGLLCQLFGGKAKKIPSGASFRGDINVLLVGDPGTSKSQLLQYVHKIAPRGIYTSGR 567
              |||:||   ||||: |          |||||||| ||||:|||  |:|: |:: | |:|:|:
Db        495 DIKRGLALALFGGEPKNPGGKHKVRGDINVLLCGDPGTAKSQFLKYIEKVSSRAIFTTGQ 554

Qy        568 GSSAVGLTAYVTKDPETRETVLESGALVLSDRGICCIDEFDKMSDNARSMLHEVMEQQTV 627
              |:||||||||| : | :||  ||:|||||:|||:| |||||||:|  |: :|| |||::
Db        555 GASAVGLTAYVQRHPVSREWTLEAGALVLADRGVCLIDEFDKMNDQDRTSIHEAMEQQSI 614

Qy        628 SVAKGGIIASLNARTSVLACANPSGSRYNARLSVIDNIQLPPTLLSRFDLIYLMLDKPDE 687
              |::| ||: || || |:|:| |||  | ||:  |:  :|:  |     ::|||:: :: |  |
Db        615 SISKAGIVTSLQARCTVIAAANPIGGRYDPSLTFSENVDLTEPIISRFDILCVVRDTVDP 674

Qy        688 QNDRRLARHLVALHYENYEVSKQ-------------------DALDLQTLTAYITYARQ 727
               |  |||  :||  |   ::  :|:                   : |   :  |  ||  ||::
Db        675 VQDEMLARFVVGSHVRHHPSNKEEEGLANGSAAEPAMPNTYGVEPLPQEVLKKYIIYAKE 734

Qy        728 HVHPTLSDEAAEDLINGYVEMRQKGNFPGSSKKVITATPRQLESMIRISEALARMRFSEV 787
               ||| |:    : :   | ::|::     ||     |  |  | :||||::|| ||:  :
Db        735 RVHPKLNQMDQDKVAKMYSDLRKESMATGS----IPITVRHIESMIRMAEAHARIHLRDY 790

Qy        788 VEKVDAAEAVRLLDVALQQSATDHATGTIDMDLITTGVSASERIRRAN---LLAALRELI 844
               | : |   |:|:     : :|   |    ::   :  |   |      |||  || |::|:
Db        791 VIEDDVNMAIRV----MLESFIDTQKFSVMRSMRKT-FARYLSFRRDNNELLLFILKQLV 845

Qy        845 ADKIS 849
               |::::
Db        846 AEQVT 850
```

RESULT 8
US-09-538-092-319
; Sequence 319, Application US/09538092
; Patent No. 6753314
; GENERAL INFORMATION:
; APPLICANT: Giot, Loic
; APPLICANT: Mansfield, Traci A.
; TITLE OF INVENTION: Protein-Protein Complexes and Method of Using Same
; FILE REFERENCE: 15966-542
; CURRENT APPLICATION NUMBER: US/09/538,092
; CURRENT FILING DATE: 2000-03-29
; PRIOR APPLICATION NUMBER: 60/127,352
; PRIOR FILING DATE: 1999-04-01
; PRIOR APPLICATION NUMBER: 60/178,965
; PRIOR FILING DATE: 2000-02-01
; NUMBER OF SEQ ID NOS: 1387
; SOFTWARE: CuraPatSeqFormatter Version 0.9
; SEQ ID NO 319
;   LENGTH: 1017
;   TYPE: PRT
;   ORGANISM: Saccharomyces cerevisiae
;   FEATURE:
;   NAME/KEY: misc_feature
;   LOCATION: (0)...(0)
;   OTHER INFORMATION: Polypeptide Accession Number YGL201C
US-09-538-092-319

```
  Query Match             20.2%;  Score 932;  DB 2;  Length 1017;
  Best Local Similarity   30.9%;  Pred. No. 6.1e-73;
  Matches  286;  Conservative  130;  Mismatches  334;  Indels  176;  Gaps   27;

Qy         82 SSPYDAGTPGTPGTPVATPVYATPVGTPMGTPSFHRGTPQYKQRSELGSQ-GKPLHRRRR 140
Db          2 SSPFPADTPSS-----NRPSNSSPPPSSIGA--------GFGSSSGLDSQIGSRLHFPSS 48

Qy        141 SQ-----SREPGHRSPSREPSADGRPSESAEPDDTLGGEYAYVWGT----------NVNI 185
Db         49 SQPHVSNSQTGPFVNDSTQFSSQRLQTDGSATNDMEGNEPARSFKSRALNHVKKVDDVTG 108

Qy        186 PDVLRAIRRFLHN--------------YRS-----SAHDLNSKYIQIIEETVEREEDTLN 226
Db        109 EKVREAFEQFLEDFSVQSTDTGEVEKVYRAQIEFMKIYDLNTIYID-YQHLSMRENGALA 167

Qy        227 IDMSDIYDH-----DPDLYAKIVRYPLDIIPLLDT------------------------- 256
Db        168 MAISEQYYRFLAFLQKGLRRVVRKYAPELLNTSDSLKRSEGDEGQADEDEQQDDDMNGSS 227

Qy        257 ------------------ECQEVATSLLP-TFEKHIEARPFNLKASVHMRELNPSDIDKL 297
Db        228 LPRDSGSSAAPGNGTSAMATRSITTSTSPEQTERVFQISFFNLPTVHRIRDIRSEKIGSL 287

Qy        298 VSVKGMVIRCSSIIPEIKGAFFKCLVCGHSPPLVTVVKGRVE------EPTRCEKPECAA 351
Db        288 LSISGTVTRTSEVRPELYKASFTCDMC-------RAIVDNVEQSFKYTEPTFCPNPSCEN 340

Qy        352 RNAMSLIHNRCTFANKQIVRLQETPDAIPEGETPHTVSMCLYNTMVDAVKPGDRIEVTGV 411
Db        341 RAFWTLNVTRSRFLDWQKVRIQENANEIPTGSMPRTLDVILRGDSVERAKPGDRCKFTGV 400

Qy        412 -----------------FKAMAVRVGPNQRT---------LRAL------YK-TYIDC-- 436
Db        401 EIVVPDVTQLGLPGVKPSSTLDTR-GISKTTEGLNSGVTGLRSLGVRDLTYKISFLACHV 459

Qy        437 ------VHVKKSDRGRLQTEDPMEMD---KENDMYAGYHESDTSEAANEA---KIQKLKE 484
Db        460 ISIGSNIGASSPDANSNNRETELQMAANLQANNVYQD-NERDQEVFLNSLSSDEINELKE 518

Qy        485 LSKLPGIYDRLSRSLAPSIWELEDIKKGLLCQLFGGKAKKIPSGASFRGDINVLLVGDPG 544
Db        519 MVKDEHIYDKLVRSIAPAVFGHEAVKKGILLQMLGGVHKSTVEGIKLRGDINICVVGDPS 578

Qy        545 TSKSQLLQYVHKIAPRGIYTSGRGSSAVGLTAYVTKDPETRETVLESGALVLSDRGICCI 604
Db        579 TSKSQFLKYVVGFAPRSVYTSGKASSAAGLTAAVVRDEEGGDYTIEAGALMLADNGICCI 638

Qy        605 DEFDKMSDNARSMLHEVMEQQTVSVAKGGIIASLNARTSVLACANPSGSRYNARLSVIDN 664
Db        639 DEFDKMDISDQVAIHEAMEQQTISIAKAGIHATLNARTSILAAANPVGGRYNRKLSLRGN 698

Qy        665 IQLPPTLLSRFDLIYLMLDKPDEQNDRRLARHLVALHYENYEVSKQDALDLQTLTAYITY 724
Db        699 LNMTAPIMSRFDLFFVILDDCNEKIDTELASHIVDLHMKRDE-AIEPPFSAEQLRRYIKY 757

Qy        725 ARQHVHPTLSDEAAEDLINGYVEMRQKGNFPGSSKKVITATPRQLESMIRISEALARMRF 784
Db        758 ART-FKPILTKEARSYLVEKYKELR-KDDAQGFSRSSYRITVRQLESMIRLSEAIARANC 815

Qy        785 SEVVEKVDAAEAVRLLDVALQQSATDHATGTIDMDLITTGVSASERIRRANLLAALRELI 844
Db        816 VDEITPSFIAEAYDL----LRQSIIRVDVDDVEMDEEFDNIESQSHAASGN--------- 862

Qy        845 ADKISPGSSSGLKTSQLLEDIRSQSS 870
Db        863 NDDNDDGTGSGVITSEPPADIEEGQS 888
```

RESULT 9
US-09-538-092-1158
; Sequence 1158, Application US/09538092
; Patent No. 6753314
; GENERAL INFORMATION:
; APPLICANT: Giot, Loic
; APPLICANT: Mansfield, Traci A.
; TITLE OF INVENTION: Protein-Protein Complexes and Method of Using Same
; FILE REFERENCE: 15966-542
; CURRENT APPLICATION NUMBER: US/09/538,092
; CURRENT FILING DATE: 2000-03-29
; PRIOR APPLICATION NUMBER: 60/127,352
; PRIOR FILING DATE: 1999-04-01
; PRIOR APPLICATION NUMBER: 60/178,965
; PRIOR FILING DATE: 2000-02-01
; NUMBER OF SEQ ID NOS: 1387
; SOFTWARE: CuraPatSeqFormatter Version 0.9
; SEQ ID NO 1158
;   LENGTH: 895
;   TYPE: PRT
;   ORGANISM: Homo sapiens
;   FEATURE:
;   NAME/KEY: misc_feature
;   LOCATION: (0)...(0)
;   OTHER INFORMATION: Polypeptide Accession Number P49736
US-09-538-092-1158

```
  Query Match             20.0%;  Score 919.5;  DB 2;  Length 895;
  Best Local Similarity   28.7%;  Pred. No. 6.3e-72;
  Matches  256;  Conservative  156;  Mismatches  342;  Indels  139;  Gaps   23;

Qy         38 AASPVAGRRAVRQTPTSAVRRRGRETDSARRRRSRSRSLGNSVYSSPYDAGTPGTPGTPV 97
Db          2 ASSP-AQRRRGNDPLTSSPGRSSRRTDAL--TSSPGRDL------PPFEDESEGLLGT-- 50

Qy         98 ATPVYATPVGTPM---GTPSFHRGTPQYKQRSELG------------SQGKPLHRRRRS 141
Db         51 EGPLEEEEDGEELIGDGMERDYRAIPELDAYEAEGLALDDEDVEELTASQREAAERAMRQ 110

Qy        142 QSREPG------------------HRSPSREPSADGRPSESAEPD----------DTLGG 173
Db        111 RDREAGRGLGRMRRGLLYDSDEEDEERPARKRRQVERATEDGEEDEEMIESIENLEDLKG 170

Qy        174 EYAYVW----GTNVNIPDVLRAIRRFLHNY-RSSAHDLNSKYIQIIEETVEREEDTLNID 228
Db        171 HSVREWVSMAGPRL---EIHHRFKNFLRTHVDSHGHNV---FKERISDMCKENRESLVVN 224

Qy        229 MSDIYDHDPDLYAKIVRYPLDIIPLLDTECQEVATSLLPTFEK---HIEARPFNLKASVH 285
Db        225 YEDLAAREHVLAYFLPEAPAELLQIFDEAALEVVLAMYPKYDRITNHIHVRISHLPLVEE 284

Qy        286 MRELNPSDIDKLVSVKGMVIRCSSIIPEIKGAFFKCLVCGHSPPLVTVVKGRVEEPTRCE 345
Db        285 LRSLRQLHLNQLIRTSGVVTSCTGVLPQLSMVKYNCNKCNFVLGPFCQSQNQEVKPGSC- 343

Qy        346 KPECAARNAMSLIHNRCTFANKQIVRLQETPDAIPEGETPHTVSMCLYNTMVDAVKPGDR 405
Db        344 -PECQSAGPFEVNMEETIYQNYQRIRIQESPGKVAAGRLPRSKDAILLADLVDSCNAGDE 402

Qy        406 IEVTGVFKAMAVRVGPNQRTLRA--LYKTYIDCVHVKKSDR----GRLQTEDPMEMDKEN 459
Db        403 IELTGIYHN---NYDGSLNTANGFPVFATVILANHVAKKDNKVAVGELTDED-------- 451

Qy        460 DMYAGYHESDTSEAANEAKIQKLKELSKLPGIYDRLSRSLAPSIWELEDIKKGLLCQLFG 519
Db        452 ------------------VKMITSLSKDQQIGEKIFASIAPSIYGHEDIKRGLALALFG 492

Qy        520 GKAKKIPSGASFRGDINVLLVGDPGTSKSQLLQYVHKIAPRGIYTSGRGSSAVGLTAYVT 579
Db        493 GEPKNPGGKHKVRGDINVLLCGDPGTAKSQFLKYIEKVSSRAIFTTGQGASAVGLTAYVQ 552

Qy        580 KDPETRETVLESGALVLSDRGICCIDEFDKMSDNARSMLHEVMEQQTVSVAKGGIIASLN 639
Db        553 RHPVSREWTLEAGALVLADRGVCLIDEFDKMNDQDRTSIHEAMEQQSISISKAGIVTSLQ 612

Qy        640 ARTSVLACANPSGSRYNARLSVIDNIQLPPTLLSRFDLIYLMLDKPDEQNDRRLARHLVA 699
```

```
Db        613 ARCTVIAAANPIGGRYDPSLTFSENVDLTEPIISRFDILCVVRDTVDPVQDEMLARFVVG 672

Qy        700 LHYENYEVSKQ-------------------DALDLQTLTAYITYARQHVHPTLSDEAAE 739

Db        673 SHVRHHPSNKEEEGLANGSAAEPAMPNTYGVEPLPQEVLKKYIIYAKERVHPKLNQMDQD 732

Qy        740 DLINGYVEMRQKGNFPGSSKKVITATPRQLESMIRISEALARMRFSEVVEKVDAAEAVRL 799

Db        733 KVAKMYSDLRKESMATGS----IPITVRHIESMIRMAEAHARIHLRDYVIEDDVNMAIRV 788

Qy        800 LDVALQQSATDHATGTIDMDLITTGVSASERIRRAN---LLAALRELIADKIS 849

Db        789 ----MLESFIDTQKFSVMRSMRKT-FARYLSFRRDNNELLLFILKQLVAEQVT 836
```

RESULT 10
US-09-949-016-8339
; Sequence 8339, Application US/09949016
; Patent No. 6812339
; GENERAL INFORMATION:
; APPLICANT: VENTER, J. Craig et al.
; TITLE OF INVENTION: POLYMORPHISMS IN KNOWN GENES ASSOCIATED
; TITLE OF INVENTION: WITH HUMAN DISEASE, METHODS OF DETECTION AND USES THEREOF
; FILE REFERENCE: CL001307
; CURRENT APPLICATION NUMBER: US/09/949,016
; CURRENT FILING DATE: 2000-04-14
; PRIOR APPLICATION NUMBER: 60/241,755
; PRIOR FILING DATE: 2000-10-20
; PRIOR APPLICATION NUMBER: 60/237,768
; PRIOR FILING DATE: 2000-10-03
; PRIOR APPLICATION NUMBER: 60/231,498
; PRIOR FILING DATE: 2000-09-08
; NUMBER OF SEQ ID NOS: 207012
; SOFTWARE: FastSEQ for Windows Version 4.0
; SEQ ID NO 8339
;   LENGTH: 823
;   TYPE: PRT
;   ORGANISM: Human
US-09-949-016-8339

```
  Query Match             19.9%;  Score 916.5;  DB 2;  Length 823;
  Best Local Similarity   32.3%;  Pred. No. 1e-71;
  Matches  252;  Conservative  114;  Mismatches  282;  Indels  133;  Gaps   22;

Qy        195 FLHNYRSSAHDLNSKYIQIIEETVEREEDTLNIDMSDIYDHDPDLYAKIVRYPLDIIPLL 254

Db         35 FLEEFQSS--DGEIKYLQLAEELIRPERNTLVVSFVDLEQFNQQLSTTIQEEFYRVYPYL 92

Qy        255 DTECQEVATSLLPTFEKHIEARPF---------NLKASVHMRELNPSDIDKLVSVKGMVI 305

Db         93 ---CR-----ALKTFVKDRKEIPLAKDFYVAFQDLPTRHKIRELTSSRIGLLTRISGQVV 144

Qy        306 RCSSIIPEIKGAFFKCLVCGHSPPLVTVVKG-----RVEEPTRCEKPECAARNAMSLIHN 360

Db        145 RTHPVHPELVSGTFLCLDC------QTVIRDVEQQFKYTQPNICRNPVCANRRRFLLDTN 198

Qy        361 RCTFANKQIVRLQETPDAIPEGETPHTVSMCLYNTMVDAVKPGDRIEVTGV--------- 411

Db        199 KSRFVDFQKVRIQETQAELPRGSIPRSLEVILRAEAVESAQAGDKCDFTGTLIVVPDVSK 258

Qy        412 FKAMAVRVGPNQRT-------------LRAL------YK-TYIDCVHVKKSDRGRLQTED 451

Db        259 LSTPGARAETNSRVSGVDGYETEGIRGLRALGVRDLSYRLVFLACC-------------- 304

Qy        452 PMEMDKENDMYAGYHESDTSEAANEAKIQ-------KLKELSKLPGIYDRLSRSLAPSIW 504

Db        305 ---VAPTNPRFGGKELRDEEQTAESIKNQMTVKEWEKVFEMSQDKNLYHNLCTSLFPTIH 361

Qy        505 ELEDIKKGLLCQLFGGKAKKIPSGASFRGDINVLLVGDPGTSKSQLLQYVHKIAPRGIYT 564

Db        362 GNDEVKRGVLLMLFGGVPKTTGEGTSLRGDINVCIVGDPSTAKSQFLKHVEEFSPRAVYT 421

Qy        565 SGRGSSAVGLTAYVTKDPETRETVLESGALVLSDRGICCIDEFDKMSDNARSMLHEVMEQ 624

Db        422 SGKASSAAGLTAAVVRDEESHEFVIEAGALMLADNGVCCIDEFDKMDVRDQVAIHEAMEQ 481

Qy        625 QTVSVAKGGIIASLNARTSVLACANPSGSRYNARLSVIDNIQLPPTLLSRFDLIYLMLDK 684

Db        482 QTISITKAGVKATLNARTSILAAANPISGHYDRSKSLKQNINLSAPIMSRFDLFFILVDE 541

Qy        685 PDEQNDRRLARHLVALHYENYEVSKQDALDLQTLTAYITYARQHVHPTLSDEAAEDLING 744

Db        542 CNEVTDYAIARRIVDLH-SRIEESIDRVYSLDDIRRYLLFARQ-FKPKISKESEDFIVEQ 599

Qy        745 YVEMRQKGNFPGSSKKVITATPRQLESMIRISEALARMRFSEVVEKVDAAEAVRLL---- 800

Db        600 YKHLRQRDG-SGVTKSSWRITVRQLESMIRLSEAMARMHCCDEVQPKHVKEAFRLLNKSI 658

Qy        801 ------DVALQQS-----ATDHATGTI----DMDLITTGVSA----------------- 827

Db        659 IRVETPDVNLDQEEEIQMEVDEGAGGINGHADSPAPVNGINGYNEDINQESAPKASLRLG 718

Qy        828 -SERIRRANLLAA-LRELIADKISPGSSSGLKTSQL----LEDIRSQSSVDVSLQDIKNA 881

Db        719 FSEYCRISNLIVLHLRKVEEEE----DESALKRSELVNWYLKEIESEIDSEEELINKKRI 774

Qy        882 L 882

Db        775 I 775
```

RESULT 11
US-09-248-796A-18637
; Sequence 18637, Application US/09248796A
; Patent No. 6747137
; GENERAL INFORMATION:
; APPLICANT: Keith Weinstock et al
; TITLE OF INVENTION: NUCLEIC ACID AND AMINO ACID SEQUENCES RELATING TO CANDIDA ALBICAN
; TITLE OF INVENTION: FOR DIAGNOSTICS AND THERAPEUTICS
; FILE REFERENCE: 107196.132
; CURRENT APPLICATION NUMBER: US/09/248,796A
; CURRENT FILING DATE: 1999-02-12
; PRIOR APPLICATION NUMBER: US 60/074,725
; PRIOR FILING DATE: 1998-02-13
; PRIOR APPLICATION NUMBER: US 60/096,409
; PRIOR FILING DATE: 1998-08-13
; NUMBER OF SEQ ID NOS: 28208
; SEQ ID NO 18637
;   LENGTH: 544
;   TYPE: PRT
;   ORGANISM: Candida albicans
US-09-248-796A-18637

```
  Query Match             17.7%;  Score 817.5;  DB 2;  Length 544;
  Best Local Similarity   32.2%;  Pred. No. 2.9e-63;
  Matches  188;  Conservative  108;  Mismatches  181;  Indels  107;  Gaps    9;

Qy        308 SSIIPEIKGAFFKCLVCGHSPPLVTVVKGRVEEPTRCEKPECAARNAMSLIHNRCTFANK 367

Db          1 TGVFPQLKYVKFDCLKCG--VVLVPYVQDSNTEVKISFCTNCQSKGPFKLNSEKTLYRNY 58

Qy        368 QIVRLQETPDAIPEGETPHTVSMCLYNTMVDAVKPGDRIEVTGVFKAMAVRVGPNQRTLR 427

Db         59 QRITLQEAPGTVPAGRLPRHREVILLSDLVDVAKPGEDIEVTGIYK------NNYDGNLN 112

Qy        428 A-----LYKTYIDCVHVKKSDRGRLQTEDPMEMDKENDMYAGYHESDTSEAANEAKIQKL 482

Db        113 AKNGFPVFATILEANSIRR---------------KESSAFMG--GNNLVNMWTEEEIREF 155

Qy        483 KELSKLPGIYDRLSRSLAPSIWELEDIKKGLLCQLFGGKAKKIPSGASFRGDINVLLVGD 542

Db        156 RKLSHEKGIIDKIIASMAPSIYGHKDIKTAIACSLFGGVPKDVNGKLSIRGDINVLLLGD 215

Qy        543 PGTSKSQLLQYVHKIAPRGIYTSGRGSSAVGLTAYVTKDPETRETVLESGALVLSDRGIC 602

Db        216 PGTAKSQILKYAEKTASRAVFATGQGASAVGLTASVRKDPITREWTLQGGALVLADKGTC 275
```

```
Qy        603 CIDEFDKMSDNARSMLHEVMEQQTVSVAKGGIIASLNARTSVLACANPSGSRYNARLSVI 662
Db        276 LIDEFDKMNDQDRTSIHEAMEQQSISISKAGIVTTLHARCAVIAAANPNGGRYNSTLPLP 335

Qy        663 DNIQLPPTLLSRFDLIYLMLDKPDEQNDRRLARHLVALHYENYEVSKQDALDL------- 715
Db        336 ENVDLTYPILSRFDIMCIVRDLVNPESDERLASFVIDSHMRSHPTNEEDILNANGKGGSD 395

Qy        716 -----------------------------------------QTLTAYITYARQHVHPT 732
Db        396 AQDDDENMEDEDGVDQPSAARTRSERIEQLNKQKEQEISPIPQDLLIKYIQYARVKIQPK 455

Qy        733 LSDEAAEDLINGYVEMRQK----GNFPGSSKKVITATPRQLESMIRISEALARMRFSEVV 788
Db        456 LHQMNMDKLARVYADLRKEAITTGSYP--------ITVRHLESILRIAESFAKMRLSEFV 507

Qy        789 EKVDAAEAVRLLDVALQQSATDHATGTIDMDLITTGVSASERIR 832
Db        508 SQNDLNRAIKV---------------SIDSFIGAQKVTVKQKLR 536
```

RESULT 12
US-09-248-796A-20007
; Sequence 20007, Application US/09248796A
; Patent No. 6747137
; GENERAL INFORMATION:
; APPLICANT: Keith Weinstock et al
; TITLE OF INVENTION: NUCLEIC ACID AND AMINO ACID SEQUENCES RELATING TO CANDIDA ALBICAN
; TITLE OF INVENTION: FOR DIAGNOSTICS AND THERAPEUTICS
; FILE REFERENCE: 107196.132
; CURRENT APPLICATION NUMBER: US/09/248,796A
; CURRENT FILING DATE: 1999-02-12
; PRIOR APPLICATION NUMBER: US 60/074,725
; PRIOR FILING DATE: 1998-02-13
; PRIOR APPLICATION NUMBER: US 60/096,409
; PRIOR FILING DATE: 1998-08-13
; NUMBER OF SEQ ID NOS: 28208
; SEQ ID NO 20007
;   LENGTH: 495
;   TYPE: PRT
;   ORGANISM: Candida albicans
;   FEATURE:
;   NAME/KEY: UNSURE
;   LOCATION: (1)
;   OTHER INFORMATION: Identity of amino acid sequences at the above locations are unkno
US-09-248-796A-20007

```
  Query Match             15.9%;  Score 734.5;  DB 2;  Length 495;
  Best Local Similarity   37.2%;  Pred. No. 5.5e-56;
  Matches  184;  Conservative   88;  Mismatches  170;  Indels   53;  Gaps   13;

Qy        408 VTGVFKAMAVRVGPNQRTLRALYKTYIDCVHVKKSDRGRLQTEDPMEMD-KENDMYAGYH 466
Db         11 VTGL-KSLGVR--------DLTYKLAFGACHV-ASMVNKAGGNEQLEVDLNDQEVFL--- 57

Qy        467 ESDTSEAANEAKIQKLKELSKLPGIYDRLSRSLAPSIWELEDIKKGLLCQLFGGKAKKIP 526
Db         58 -----TSLSDAEVLQLKEMVKDEHIYDKLVNSIAPAVFGHEVIKKGILLQLLGGVHKQTV 112

Qy        527 SGASFRGDINVLLVGDPGTSKSQLLQYVHKIAPRGIYTSGRGSSAVGLTAYVTKDPETRE 586
Db        113 DGINLRGDINICIVGDPSTSKSQFLKYVCGFSPRAVYTSGKASSAAGLTAAVVKDEESGE 172

Qy        587 TVLESGALVLSDRGICCIDEFDKMSDNARSMLHEVMEQQTVSVAKGGIIASLNARTSVLA 646
Db        173 YTIEAGALMLADNGICAIDEFDKMDIADQVAIHEAMEQQTISIAKAGIHATLNARTSILA 232

Qy        647 CANPSGSRYNARLSVIDNIQLPPTLLSRFDLIYLMLDKPDEQNDRRLARHLVALHYENYE 706
Db        233 AANPIGGRYNRKLGLRSNLNMTAPIMSRFDLFFVVLDDCNERIDTQLASHIVDLH----- 287

Qy        707 VSKQDALD----LQTLTAYITYARQHVHPTLSDEAAEDLINGYVEMRQKGNFPGSSKKVI 762
Db        288 MLRDDAIDPPYSAEQLARYIKYAKT-FKPRMTKEARDFLVTRYKELRE-DDAQGLGRSSY 345

Qy        763 TATPRQLESMIRISEALARMRFSEVVEKVDAAEAVRLLDVA--------LQQSATDHATG 814
Db        346 RITVRQLESMIRLSEAIARANCTEEITPSFVAEAYDLLKQSIIRVEMDDIEMDDEERPTG 405

Qy        815 --------TIDMDLITTGVSASERIRRANL----LAALRELIADKIS---PGSSSGLKTS 859
Db        406 EPGQAQEQDQPQDQDTQQSQPSERVPPVSISYDKYVSIMNMLVKKITEDEKNGGDGLSAD 465

Qy        860 QLLEDIRSQSSVDVS 874
Db        466 TLVEWYLLQKEDEIS 480
```

RESULT 13
US-09-248-796A-18673
; Sequence 18673, Application US/09248796A
; Patent No. 6747137
; GENERAL INFORMATION:
; APPLICANT: Keith Weinstock et al
; TITLE OF INVENTION: NUCLEIC ACID AND AMINO ACID SEQUENCES RELATING TO CANDIDA ALBICAN
; TITLE OF INVENTION: FOR DIAGNOSTICS AND THERAPEUTICS
; FILE REFERENCE: 107196.132
; CURRENT APPLICATION NUMBER: US/09/248,796A
; CURRENT FILING DATE: 1999-02-12
; PRIOR APPLICATION NUMBER: US 60/074,725
; PRIOR FILING DATE: 1998-02-13
; PRIOR APPLICATION NUMBER: US 60/096,409
; PRIOR FILING DATE: 1998-08-13
; NUMBER OF SEQ ID NOS: 28208
; SEQ ID NO 18673
;   LENGTH: 331
;   TYPE: PRT
;   ORGANISM: Candida albicans
US-09-248-796A-18673

```
  Query Match             14.4%;  Score 662;  DB 2;  Length 331;
  Best Local Similarity   41.5%;  Pred. No. 7.3e-50;
  Matches  151;  Conservative   60;  Mismatches  109;  Indels   44;  Gaps    7;

Qy        318 FFKCLVCGHSPPLVTVVKGRVEEPTRCEKPECAARNAMSLIHNRCTFANK----QIVRLQ 373
Db          6 FSKKSITKHSLPL-----------TECNSPSGVNDNNKGQLF-MSTRASKLSAFQEVKIQ 53

Qy        374 ETPDAIPEGETPHTVSMCLYNTMVDAVKPGDRIEVTGVFKAMAVRVGPNQRTLRA--LYK 431
Db         54 ELSSQVPVGHIPRSLTVHVNGDLVRSMNPGDTVDLSGIFMPSPY---TGYRALKAGLLTE 110

Qy        432 TYIDCVHVKKSDRGRLQTEDPMEMDKENDMYAGYHESDTSEAANEAKIQKLKELSKLPGI 491
Db        111 TYLEAQHVKQHKK---------------------QYDSMTLSSQAQ-DKIDELLLQGDV 147

Qy        492 YDRLSRSLAPSIWELEDIKKGLLCQLFGGKAKKIPSGASFRGDINVLLVGDPGTSKSQLL 551
Db        148 YNKLAKSIAPEIYGHLDVKKILLLLLCGGVTKEIGDGLKIRGDINVCLMGDPGVAKSQLL 207

Qy        552 QYVHKIAPRGIYTSGRGSSAVGLTAYVTKDPETRETVLESGALVLSDRGICCIDEFDKMS 611
Db        208 KAIGKIAPRSVYTTGRGSSGVGLTAAVMRDPITDEMVLEGGALVLADNGICCIDEFDKMD 267

Qy        612 DNARSMLHEVMEQQTVSVAKGGIIASLNARTSVLACANPSGSRYNARLSVIDNIQLPPTL 671
Db        268 ESDRTAIHEVMEQQTISIAKAGITTTLNARTSILAAANPLYGRYNPRLSPHENINLPASF 327

Qy        672 LSRF 675
Db        328 VVPF 331
```

RESULT 14
US-09-248-796A-18814

```
; Sequence 18814, Application US/09248796A
; Patent No. 6747137
; GENERAL INFORMATION:
;  APPLICANT: Keith Weinstock et al
;  TITLE OF INVENTION: NUCLEIC ACID AND AMINO ACID SEQUENCES RELATING TO CANDIDA ALBICAN
;  TITLE OF INVENTION:  FOR DIAGNOSTICS AND THERAPEUTICS
;  FILE REFERENCE: 107196.132
;  CURRENT APPLICATION NUMBER: US/09/248,796A
;  CURRENT FILING DATE:  1999-02-12
;  PRIOR APPLICATION NUMBER: US 60/074,725
;  PRIOR FILING DATE: 1998-02-13
;  PRIOR APPLICATION NUMBER: US 60/096,409
;  PRIOR FILING DATE: 1998-08-13
;  NUMBER OF SEQ ID NOS: 28208
; SEQ ID NO 18814
;   LENGTH: 435
;   TYPE: PRT
;   ORGANISM: Candida albicans
US-09-248-796A-18814

  Query Match             10.3%;  Score 474;  DB 2;  Length 435;
  Best Local Similarity   28.5%;  Pred. No. 5.2e-33;
  Matches  135;  Conservative   84;  Mismatches  170;  Indels   84;  Gaps   14;

Qy        161 PSESAEPDDTLGGEYAYVWGTNVNIP------DVLRAIRRFLHNYRSSAHDLNSKYIQII 214
              | :|: |||                 ||       | :|  : ||     |:        |  ||
Db         10 PPPTADQDDT------------NQPLDAIFGDRVRRFQEFLDRIDSNT---GIDYRSII 53

Qy        215 EETVEREEDTLNIDMSDIYDHDPDLYAKIVRYPLDIIPLLDTECQEVATSLL----PTFE 270
              :: : : :  |:: : :| : | : :  ::  | | :|  :   ::   ::    |:|
Db         54 KDMLIKSKFRLSVSIDEIREFDREFWLGLLNQPADYLPACERALRDTVLAIYDPQDPSF- 112

Qy        271 KHIEARP-----FNLKASVHMRELNPSDID-----KLVSVKGMVIRCSSIIPEI------ 314
               |     |        : | :       | |  ||       |:||::|:| | | : |::
Db        113 PHDSYDPNQQYYLSFKGAFGGHSLTPRSIDSSYLSKMVSIEGIVTRASLVRPKVIRSVHY 172

Qy        315 ---KGAFFKCLVCGHSPPLVTVVKGRVEEPTRCEKPECAARNAMSLIHNRCTFANKQIVR 371
                 | |:       :         :  ||        |       | :: :     |: : | :
Db        173 AEKTGRFYAREYRDQTTSFDAIATPAI-YPT-----EDMEGNKLTTEYGYSTYRDYQKIS 226

Qy        372 LQETPDAIPEGETPHTVSMCLYNTMVDAVKPGDRIEVTGVFKAMAVRVGPNQRTLRALYK 431
              :|| |:  | |:  | :|  : | : :||  |||||::: ||::|:         :    : :|
Db        227 VQEMPETAPPGQLPRSVDVILDDDLVDLTKPGDRVQIVGVYRALGGAANNS-----SSFK 281

Qy        432 TYIDCVHVKKSDRGRLQTEDPMEMDKENDMYAGYHESDTSEAANE----AKIQKLKELSK 487
              | |                        | :|    |     |   |:  |         |: : :|:|
Db        282 TVI---------------------LSNSVYL-LHARSTGVASQEKLTDQDIRNINKLAK 318

Qy        488 LPGIYDRLSRSLAPSIWELEDIKKGLLCQLFGGKAKKIPSGASFRGDINVLLVGDPGTSK 547
                |:| ||||||||:  : ||| :|   : ||   | : :|    ||||:|:|||| |:|
Db        319 DRKIFDILSRSLAPSIYGFDYIKKAVLLMMGGVEKNLDNGTHLRGDINILMVGDPSTAK 378

Qy        548 SQLLQYVHKIAPRGIYTSGRGSSAVGLTAYVTKDPETRETVLESGALVLSDRG 600
              ||:|::|    |     | |:|||| ||||  || | ||  |   ||:||::     |
Db        379 SQVLRFVLNTASLAIATTGRGSSGVGLTAAVTTDKETGERRLEAGAILTRRAG 431

RESULT 15
US-09-248-796A-20019
; Sequence 20019, Application US/09248796A
; Patent No. 6747137
; GENERAL INFORMATION:
;  APPLICANT: Keith Weinstock et al
;  TITLE OF INVENTION: NUCLEIC ACID AND AMINO ACID SEQUENCES RELATING TO CANDIDA ALBICAN
;  TITLE OF INVENTION:  FOR DIAGNOSTICS AND THERAPEUTICS
;  FILE REFERENCE: 107196.132
;  CURRENT APPLICATION NUMBER: US/09/248,796A
;  CURRENT FILING DATE:  1999-02-12
;  PRIOR APPLICATION NUMBER: US 60/074,725
;  PRIOR FILING DATE: 1998-02-13
;  PRIOR APPLICATION NUMBER: US 60/096,409
;  PRIOR FILING DATE: 1998-08-13
;  NUMBER OF SEQ ID NOS: 28208
; SEQ ID NO 20019
;   LENGTH: 173
;   TYPE: PRT
;   ORGANISM: Candida albicans
US-09-248-796A-20019

  Query Match              8.5%;  Score 391.5;  DB 2;  Length 173;
  Best Local Similarity   47.0%;  Pred. No. 2.2e-26;
  Matches   70;  Conservative   40;  Mismatches   36;  Indels    3;  Gaps    3;

Qy        262 ATSLLPTFEKHI-EARPFNLK-ASVHMRELNPSDIDKLVSVKGMVIRCSSIIPEIKGAFF 319
              | | |     | ::    ||:|:          :|||||:|||||||||||: :| :||||::| |||
Db         17 AQSKLDEIETNVYTIRPYNVNLVEKGIRELNPNDIDKLVSVKGLTLRSTSIIPDMKVAFF 76

Qy        320 KCLVCGHSPPLVTVVKGRVEEPTRCEKPECAARNAMSLIHNRCTFANKQIVRLQETPDAI 379
              :|  |||:    | : :| : |||:| :   |     |:| ||||| :|::||:::||||| :
Db         77 RCNACGHTVG-VEIDRGVISEPTKCPREVCGQTNSMVLIHNRSSFSDKQVIKLQETPDLV 135

Qy        380 PEGETPHTVSMCLYNTMVDAVKPGDRIEV 408
              |:|:|||::::|:|: :||: : |||:::
Db        136 PDGQTPHSINLCVYDELVDSCRAGDRVDL 164

Search completed: December  7, 2005, 17:44:36
Job time : 50 secs
```

OM protein - protein search, using sw model

Run on: December 7, 2005, 17:35:36 ; Search time 169 Seconds (without alignments) 2227.599 Million cell updates/sec

Title: US-10-768-511-8
Perfect score: 4609
Sequence: 1 MENNDALDIGAVSSPYPSQS..........LGSLQGEGFLTVHGDIVKRV 901

Scoring table: BLOSUM62
Gapop 10.0 , Gapext 0.5

Searched: 1867569 seqs, 417829326 residues

Total number of hits satisfying chosen parameters: 1867569

Minimum DB seq length: 0
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries

Database : Published_Applications_AA_Main:*
1: /cgn2_6/ptodata/1/pubpaa/US07_PUBCOMB.pep:*
2: /cgn2_6/ptodata/1/pubpaa/US08_PUBCOMB.pep:*
3: /cgn2_6/ptodata/1/pubpaa/US09_PUBCOMB.pep:*
4: /cgn2_6/ptodata/1/pubpaa/US10A_PUBCOMB.pep:*
5: /cgn2_6/ptodata/1/pubpaa/US10B_PUBCOMB.pep:*
6: /cgn2_6/ptodata/1/pubpaa/US11_PUBCOMB.pep:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	% Query Match	Length	DB	ID	Description
1	4609	100.0	901	3	US-09-828-062-8	Sequence 8, Appli
2	4609	100.0	901	4	US-10-768-511-8	Sequence 8, Appli
3	2550	55.3	881	4	US-10-425-114-65404	Sequence 65404, A
4	2474.5	53.7	1017	4	US-10-437-963-147122	Sequence 147122,
5	1972.5	42.8	551	4	US-10-425-115-283561	Sequence 283561,
6	1887.5	41.0	851	4	US-10-425-115-283560	Sequence 283560,
7	1883	40.9	1023	4	US-10-128-714-8558	Sequence 8558, Ap
8	1860	40.4	874	4	US-10-128-714-3558	Sequence 3558, Ap
9	1835.5	39.8	989	4	US-10-320-797-3332	Sequence 3332, Ap
10	1835	39.8	960	4	US-10-369-493-3801	Sequence 3801, Ap
11	1833	39.8	933	4	US-10-369-493-22511	Sequence 22511, A
12	1817.5	39.4	923	5	US-10-756-149-5246	Sequence 5246, Ap
13	1754	38.1	910	4	US-10-032-585-7480	Sequence 7480, Ap
14	1686.5	36.6	866	6	US-11-097-143-1689	Sequence 1689, Ap
15	1670.5	36.2	456	4	US-10-425-114-68963	Sequence 68963, A
16	1500	32.5	712	5	US-10-723-860-2090	Sequence 2090, Ap
17	1500	32.5	712	5	US-10-756-149-5255	Sequence 5255, Ap
18	1375	29.8	724	3	US-09-925-300-1053	Sequence 1053, Ap
19	1020.5	22.1	666	4	US-10-369-493-21538	Sequence 21538, A
20	1016.5	22.1	760	4	US-10-369-493-22564	Sequence 22564, A
21	1003	21.8	814	4	US-10-128-714-8100	Sequence 8100, Ap
22	997.5	21.6	287	4	US-10-425-114-39678	Sequence 39678, A
23	985.5	21.4	718	4	US-10-128-714-3399	Sequence 3399, Ap
24	983.5	21.3	713	4	US-10-128-714-8399	Sequence 8399, Ap
25	983.5	21.3	809	4	US-10-032-585-7209	Sequence 7209, Ap
26	976.5	21.2	720	6	US-11-097-143-10251	Sequence 10251, A
27	973.5	21.1	632	4	US-10-128-714-3100	Sequence 3100, Ap
28	973	21.1	699	4	US-10-369-493-22832	Sequence 22832, A
29	972	21.1	743	4	US-10-425-114-58624	Sequence 58624, A
30	970	21.0	766	3	US-09-925-301-1382	Sequence 1382, Ap
31	968.5	21.0	720	4	US-10-425-115-258341	Sequence 258341,
32	963.5	20.9	720	4	US-10-101-080-2	Sequence 2, Appli
33	963	20.9	775	4	US-10-369-493-1800	Sequence 1800, Ap
34	961.5	20.9	881	4	US-10-369-493-5601	Sequence 5601, Ap
35	960.5	20.8	719	5	US-10-733-878-461	Sequence 461, App
36	958	20.8	744	4	US-10-320-797-3279	Sequence 3279, Ap
37	956.5	20.8	698	4	US-10-369-493-18198	Sequence 18198, A
38	951	20.6	746	4	US-10-369-493-4174	Sequence 4174, Ap
39	949.5	20.6	733	6	US-11-097-143-7995	Sequence 7995, Ap
40	948	20.6	757	4	US-10-320-797-3296	Sequence 3296, Ap
41	947.5	20.6	716	5	US-10-739-930-6189	Sequence 6189, Ap
42	946.5	20.5	904	3	US-09-964-899-49	Sequence 49, Appl
43	946.5	20.5	904	5	US-10-975-523-49	Sequence 49, Appl
44	944	20.5	546	4	US-10-369-493-22683	Sequence 22683, A
45	943.5	20.5	592	4	US-10-369-493-11228	Sequence 11228, A

ALIGNMENTS

```
RESULT 1
US-09-828-062-8
; Sequence 8, Application US/09828062
; Publication No. US20030097675A1
; GENERAL INFORMATION:
;  APPLICANT: COSTA E SILVA, OSWALDO DA
;  APPLICANT:  BOHNERT, HANS J.
;  APPLICANT:  VAN THIELEN, NOCHA
;  APPLICANT:  CHEN, ROUYING
;  APPLICANT:  SARRIA-MILLAN, RODRIGO
;  TITLE OF INVENTION: CELL CYCLE STRESS-RELATED PROTEINS AND METHODS OF USE
;  TITLE OF INVENTION:  IN PLANTS
;  FILE REFERENCE: 16313-0031
;  CURRENT APPLICATION NUMBER: US/09/828,062
;  CURRENT FILING DATE:  2001-08-20
;  PRIOR APPLICATION NUMBER: 60/196,001
;  PRIOR FILING DATE: 2000-04-07
;  NUMBER OF SEQ ID NOS: 34
;  SOFTWARE: PatentIn Ver. 2.1
; SEQ ID NO 8
;   LENGTH: 901
;   TYPE: PRT
;   ORGANISM: Physcomitrella patens
US-09-828-062-8

  Query Match             100.0%;  Score 4609;  DB 3;  Length 901;
  Best Local Similarity   100.0%;  Pred. No. 0;
  Matches  901;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy          1 MENNDALDIGAVSSPYPSQSEGVSTPLPQVTSPSFDNAASPVAGRRAVRQTPTSAVRRRG 60
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db          1 MENNDALDIGAVSSPYPSQSEGVSTPLPQVTSPSFDNAASPVAGRRAVRQTPTSAVRRRG 60

Qy         61 RETDSARRRRSRSRSLGNSVYSSPYDAGTPGTPGTPVATPVYATPVGTPMGTPSFHRGTP 120
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db         61 RETDSARRRRSRSRSLGNSVYSSPYDAGTPGTPGTPVATPVYATPVGTPMGTPSFHRGTP 120

Qy        121 QYKQRSELGSQGKPLHRRRRSQSREPGHRSPSREPSADGRPSESAEPDDTLGGEYAYVWG 180
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        121 QYKQRSELGSQGKPLHRRRRSQSREPGHRSPSREPSADGRPSESAEPDDTLGGEYAYVWG 180

Qy        181 TNVNIPDVLRAIRRFLHNYRSSAHDLNSKYIQIIEETVEREEDTLNIDMSDIYDHDPDLY 240
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        181 TNVNIPDVLRAIRRFLHNYRSSAHDLNSKYIQIIEETVEREEDTLNIDMSDIYDHDPDLY 240

Qy        241 AKIVRYPLDIIPLLDTECQEVATSLLPTFEKHIEARPFNLKASVHMRELNPSDIDKLVSV 300
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        241 AKIVRYPLDIIPLLDTECQEVATSLLPTFEKHIEARPFNLKASVHMRELNPSDIDKLVSV 300

Qy        301 KGMVIRCSSIIPEIKGAFFKCLVCGHSPPLVTVVKGRVEEPTRCEKPECAARNAMSLIHN 360
```

```
                 ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db          301 KGMVIRCSSIIPEIKGAFFKCLVCGHSPPLVTVVKGRVEEPTRCEKPECAARNAMSLIHN 360

Qy          361 RCTFANKQIVRLQETPDAIPEGETPHTVSMCLYNTMVDAVKPGDRIEVTGVFKAMAVRVG 420
                ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db          361 RCTFANKQIVRLQETPDAIPEGETPHTVSMCLYNTMVDAVKPGDRIEVTGVFKAMAVRVG 420

Qy          421 PNQRTLRALYKTYIDCVHVKKSDRGRLQTEDPMEMDKENDMYAGYHESDTSEAANEAKIQ 480
                ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db          421 PNQRTLRALYKTYIDCVHVKKSDRGRLQTEDPMEMDKENDMYAGYHESDTSEAANEAKIQ 480

Qy          481 KLKELSKLPGIYDRLSRSLAPSIWELEDIKKGLLCQLFGGKAKKIPSGASFRGDINVLLV 540
                ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db          481 KLKELSKLPGIYDRLSRSLAPSIWELEDIKKGLLCQLFGGKAKKIPSGASFRGDINVLLV 540

Qy          541 GDPGTSKSQLLQYVHKIAPRGIYTSGRGSSAVGLTAYVTKDPETRETVLESGALVLSDRG 600
                ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db          541 GDPGTSKSQLLQYVHKIAPRGIYTSGRGSSAVGLTAYVTKDPETRETVLESGALVLSDRG 600

Qy          601 ICCIDEFDKMSDNARSMLHEVMEQQTVSVAKGGIIASLNARTSVLACANPSGSRYNARLS 660
                ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db          601 ICCIDEFDKMSDNARSMLHEVMEQQTVSVAKGGIIASLNARTSVLACANPSGSRYNARLS 660

Qy          661 VIDNIQLPPTLLSRFDLIYLMLDKPDEQNDRRLARHLVALHYENYEVSKQDALDLQTLTA 720
                ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db          661 VIDNIQLPPTLLSRFDLIYLMLDKPDEQNDRRLARHLVALHYENYEVSKQDALDLQTLTA 720

Qy          721 YITYARQHVHPTLSDEAAEDLINGYVEMRQKGNFPGSSKKVITATPRQLESMIRISEALA 780
                ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db          721 YITYARQHVHPTLSDEAAEDLINGYVEMRQKGNFPGSSKKVITATPRQLESMIRISEALA 780

Qy          781 RMRFSEVVEKVDAAEAVRLLDVALQQSATDHATGTIDMDLITTGVSASERIRRANLLAAL 840
                ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db          781 RMRFSEVVEKVDAAEAVRLLDVALQQSATDHATGTIDMDLITTGVSASERIRRANLLAAL 840

Qy          841 RELIADKISPGSSSGLKTSQLLEDIRSQSSVDVSLQDIKNALGSLQGEGFLTVHGDIVKR 900
                ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db          841 RELIADKISPGSSSGLKTSQLLEDIRSQSSVDVSLQDIKNALGSLQGEGFLTVHGDIVKR 900

Qy          901 V 901
                |
Db          901 V 901
```

RESULT 2
US-10-768-511-8
; Sequence 8, Application US/10768511
; Publication No. US20040128721A1
; GENERAL INFORMATION:
; APPLICANT: COSTA E SILVA, OSWALDO DA
; APPLICANT: BOHNERT, HANS J.
; APPLICANT: VAN THIELEN, NOCHA
; APPLICANT: CHEN, ROUYING
; APPLICANT: SARRIA-MILLAN, RODRIGO
; TITLE OF INVENTION: CELL CYCLE STRESS-RELATED PROTEINS AND METHODS OF USE
; TITLE OF INVENTION: IN PLANTS
; FILE REFERENCE: 16313-0031
; CURRENT APPLICATION NUMBER: US/10/768,511
; CURRENT FILING DATE: 2004-01-30
; PRIOR APPLICATION NUMBER: 60/196,001
; PRIOR FILING DATE: 2000-04-07
; NUMBER OF SEQ ID NOS: 34
; SOFTWARE: PatentIn Ver. 2.1
; SEQ ID NO 8
; LENGTH: 901
; TYPE: PRT
; ORGANISM: Physcomitrella patens
US-10-768-511-8

```
  Query Match             100.0%;  Score 4609;  DB 4;  Length 901;
  Best Local Similarity   100.0%;  Pred. No. 0;
  Matches  901;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy            1 MENNDALDIGAVSSPYPSQSEGVSTPLPQVTSPSFDNAASPVAGRRAVRQTPTSAVRRRG 60
                ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db            1 MENNDALDIGAVSSPYPSQSEGVSTPLPQVTSPSFDNAASPVAGRRAVRQTPTSAVRRRG 60

Qy           61 RETDSARRRRSRSRSLGNSVYSSPYDAGTPGTPGTPVATPVYATPVGTPMGTPSFHRGTP 120
                ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db           61 RETDSARRRRSRSRSLGNSVYSSPYDAGTPGTPGTPVATPVYATPVGTPMGTPSFHRGTP 120

Qy          121 QYKQRSELGSQGKPLHRRRRSQSREPGHRSPSREPSADGRPSESAEPDDTLGGEYAYVWG 180
                ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db          121 QYKQRSELGSQGKPLHRRRRSQSREPGHRSPSREPSADGRPSESAEPDDTLGGEYAYVWG 180

Qy          181 TNVNIPDVLRAIRRFLHNYRSSAHDLNSKYIQIIEETVEREEDTLNIDMSDIYDHDPDLY 240
                ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db          181 TNVNIPDVLRAIRRFLHNYRSSAHDLNSKYIQIIEETVEREEDTLNIDMSDIYDHDPDLY 240

Qy          241 AKIVRYPLDIIPLLDTECQEVATSLLPTFEKHIEARPFNLKASVHMRELNPSDIDKLVSV 300
                ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db          241 AKIVRYPLDIIPLLDTECQEVATSLLPTFEKHIEARPFNLKASVHMRELNPSDIDKLVSV 300

Qy          301 KGMVIRCSSIIPEIKGAFFKCLVCGHSPPLVTVVKGRVEEPTRCEKPECAARNAMSLIHN 360
                ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db          301 KGMVIRCSSIIPEIKGAFFKCLVCGHSPPLVTVVKGRVEEPTRCEKPECAARNAMSLIHN 360

Qy          361 RCTFANKQIVRLQETPDAIPEGETPHTVSMCLYNTMVDAVKPGDRIEVTGVFKAMAVRVG 420
                ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db          361 RCTFANKQIVRLQETPDAIPEGETPHTVSMCLYNTMVDAVKPGDRIEVTGVFKAMAVRVG 420

Qy          421 PNQRTLRALYKTYIDCVHVKKSDRGRLQTEDPMEMDKENDMYAGYHESDTSEAANEAKIQ 480
                ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db          421 PNQRTLRALYKTYIDCVHVKKSDRGRLQTEDPMEMDKENDMYAGYHESDTSEAANEAKIQ 480

Qy          481 KLKELSKLPGIYDRLSRSLAPSIWELEDIKKGLLCQLFGGKAKKIPSGASFRGDINVLLV 540
                ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db          481 KLKELSKLPGIYDRLSRSLAPSIWELEDIKKGLLCQLFGGKAKKIPSGASFRGDINVLLV 540

Qy          541 GDPGTSKSQLLQYVHKIAPRGIYTSGRGSSAVGLTAYVTKDPETRETVLESGALVLSDRG 600
                ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db          541 GDPGTSKSQLLQYVHKIAPRGIYTSGRGSSAVGLTAYVTKDPETRETVLESGALVLSDRG 600

Qy          601 ICCIDEFDKMSDNARSMLHEVMEQQTVSVAKGGIIASLNARTSVLACANPSGSRYNARLS 660
                ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db          601 ICCIDEFDKMSDNARSMLHEVMEQQTVSVAKGGIIASLNARTSVLACANPSGSRYNARLS 660

Qy          661 VIDNIQLPPTLLSRFDLIYLMLDKPDEQNDRRLARHLVALHYENYEVSKQDALDLQTLTA 720
                ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db          661 VIDNIQLPPTLLSRFDLIYLMLDKPDEQNDRRLARHLVALHYENYEVSKQDALDLQTLTA 720

Qy          721 YITYARQHVHPTLSDEAAEDLINGYVEMRQKGNFPGSSKKVITATPRQLESMIRISEALA 780
                ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db          721 YITYARQHVHPTLSDEAAEDLINGYVEMRQKGNFPGSSKKVITATPRQLESMIRISEALA 780

Qy          781 RMRFSEVVEKVDAAEAVRLLDVALQQSATDHATGTIDMDLITTGVSASERIRRANLLAAL 840
                ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db          781 RMRFSEVVEKVDAAEAVRLLDVALQQSATDHATGTIDMDLITTGVSASERIRRANLLAAL 840

Qy          841 RELIADKISPGSSSGLKTSQLLEDIRSQSSVDVSLQDIKNALGSLQGEGFLTVHGDIVKR 900
                ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db          841 RELIADKISPGSSSGLKTSQLLEDIRSQSSVDVSLQDIKNALGSLQGEGFLTVHGDIVKR 900

Qy          901 V 901
                |
Db          901 V 901
```

RESULT 3
US-10-425-114-65404
; Sequence 65404, Application US/10425114
; Publication No. US20040034888A1
; GENERAL INFORMATION:
; APPLICANT: Liu, Jingdong
; APPLICANT: Zhou, Yihua
; APPLICANT: Kovalic, David K.

```
;  APPLICANT:  Screen, Steven E
;  APPLICANT:  Tabaska, Jack E
;  APPLICANT:  Cao, Yongwei
;  TITLE OF INVENTION: Nucleic Acid Molecules and Other Molecules Associated With
;  TITLE OF INVENTION:  Plants and Uses Thereof for Plant Improvement
;  FILE REFERENCE: 38-21(53313)B
;  CURRENT APPLICATION NUMBER: US/10/425,114
;  CURRENT FILING DATE:  2003-04-28
;  NUMBER OF SEQ ID NOS: 73128
; SEQ ID NO 65404
;   LENGTH: 881
;   TYPE: PRT
;   ORGANISM: Zea mays
;   FEATURE:
;   OTHER INFORMATION: Clone ID: LIB4763-010-C6_FLI.pep
US-10-425-114-65404

  Query Match             55.3%;  Score 2550;  DB 4;  Length 881;
  Best Local Similarity   57.3%;  Pred. No. 2.2e-182;
  Matches  514;  Conservative  143;  Mismatches  176;  Indels   64;  Gaps   17;

Qy         20 SEGVSTPLPQVTSPSFDNAASPVAGRRAVRQTPTSAVRRRGRETDSARRRRSRSRSLGNS 79
Db         34 SDGGGSSPPPASSP-YGRPSSPLA---VTNSSPSQPTRRPG-----GRRRRGS------- 77

Qy         80 VYSSPYDAGTPGTPGTPVATPVYATPVGTPMGTPSFHRGTPQYKQRSELGSQGKPLHRRR 139
Db         78 --ASPY-ASSPSLGG-------FETP-------PHPGRRTP-----SGAGA-GAPRQPRQ 114

Qy        140 RSQSREPGHRSPSREPSADG-RPSESAEPDDTLGG-----EYAYVWGTNVNIPDVLRAIR 193
Db        115 NSTGRFP--PTPSTPMSTDDVPPSSEAGDDETDGGGGVDATPVFVWGTNISVQDVNAAIL 172

Qy        194 RFLHNYRSSAHDL--------NSKYIQIIEETVERE-EDTLNIDMSDIYDHDPDLYAKIV 244
Db        173 RFLRHFR-DPRDAGRVDPVMDEGKYMRAIHRILELEGGESLDVDAHDVFDHDPDLYSKMV 231

Qy        245 RYPLDIIPLLDTECQEVATSLLPTFEKHIEARPFNLKASVHMRELNPSDIDKLVSVKGMV 304
Db        232 RYPLEVLAIFDIVLMDLVARIEPLFEKHIQTRIYNLKSSICLRNLNPSDIEKMVSIKGMI 291

Qy        305 IRCSSIIPEIKGAFFKCLVCGHSPPLVTVVKGRVEEPTRCEKPECAARNAMSLIHNRCTF 364
Db        292 IRCSSVIPELKEAVFRCLVCGFYSEPVMVDRGRVTEPHICQKEQCKATNSMTLVHNRCRF 351

Qy        365 ANKQIVRLQETPDAIPEGETPHTVSMCLYNTMVDAVKPGDRIEVTGVFKAMAVRVGPNQR 424
Db        352 SDKQIIKLQETPDEIPEGGTPHTVSVLMHDKLVDAGKPGDRVEITGIYRAMSIRIGPTQR 411

Qy        425 TLRALYKTYIDCVHVKKSDRGRLQTEDPMEMDKENDMYAGYHESDTSEAANEAKIQKLKE 484
Db        412 TVKSIFKTYIDCLHIKKTDKSRLHVEDTMDIDNSN------ASKSTEEDFLSDKVEKLKE 465

Qy        485 LSKLPGIYDRLSRSLAPSIWELEDIKKGLLCQLFGGKAKKIPSGASFRGDINVLLVGDPG 544
Db        466 LSKLPDIYERLTRSLAPNIWELDDVKRGLLCQLFGGNPLKLPSGASFRGDINILLVGDPG 525

Qy        545 TSKSQLLQYVHKIAPRGIYTSGRGSSAVGLTAYVTKDPETRETVLESGALVLSDRGICCI 604
Db        526 TSKSQLLQYMHKLSPRGIYTSGRGSSAVGLTAYVTKDPETGETVLESGALVLSDKGVCCI 585

Qy        605 DEFDKMSDNARSMLHEVMEQQTVSVAKGGIIASLNARTSVLACANPSGSRYNARLSVIDN 664
Db        586 DEFDKMSDNARSMLHEVMEQQTVSIAKAGIIASLNARTSVLACANPTESRYNPRLSVIDN 645

Qy        665 IQLPPTLLSRFDLIYLMLDKPDEQNDRRLARHLVALHYENYEVSKQDALDLQTLTAYITY 724
Db        646 IHLAPTLLSRFDLIYLILDKADEQTDRRLAKHIVSLHFENPNLEELEVLDLQTLVSYISY 705

Qy        725 ARQHVHPTLSDEAAEDLINGYVEMRQKGNFPGSSKKVITATPRQLESMIRISEALARMRF 784
Db        706 ARKYIQPQLSDEAAEELTRGYVEMRKRGNSPGSRKKVITATARQIESLIRLSEALARMRF 765

Qy        785 SEVVEKVDAAEAVRLLDVALQQSATDHATGTIDMDLITTGVSASERIRRANLLAALRELI 844
Db        766 SEVVEVRDVVEAFRLLEVAMQQSATDHATGTIDMDLIMTGISASERQRRENLVAATRNLI 825

Qy        845 ADKISPGSSSGLKTSQLLEDIRSQSSVDVSLQDIKNALGSLQGEGFLTVHGDIVKRV 901
Db        826 AEKMQLGGPS-MRMIELLEELRKQSSMEIHMHELRGALGTLMTEGAVVIHGDNVRRV 881

RESULT 4
US-10-437-963-147122
; Sequence 147122, Application US/10437963
; Publication No. US20040123343A1
; GENERAL INFORMATION:
;  APPLICANT: La Rosa, Thomas J.
;  APPLICANT:  Kovalic, David K.
;  APPLICANT:  Zhou, Yihua
;  APPLICANT:  Cao, Yongwei
;  APPLICANT:  Wu, Wei
;  APPLICANT:  Boukharov, Andrey A.
;  APPLICANT:  Barbazuk, Brad
;  APPLICANT:  Li, Ping
;  TITLE OF INVENTION: Rice Nucleic Acid Molecules and Other Molecules Associated With
;  TITLE OF INVENTION:  Plants and Uses Thereof for Plant Improvement
;  FILE REFERENCE: 38-21(53221)B
;  CURRENT APPLICATION NUMBER: US/10/437,963
;  CURRENT FILING DATE:  2003-05-14
;  NUMBER OF SEQ ID NOS: 204966
; SEQ ID NO 147122
;   LENGTH: 1017
;   TYPE: PRT
;   ORGANISM: Oryza sativa
;   FEATURE:
;   OTHER INFORMATION: Clone ID: PAT_MRT4530_47680C.1.pep
US-10-437-963-147122

  Query Match             53.7%;  Score 2474.5;  DB 4;  Length 1017;
  Best Local Similarity   56.4%;  Pred. No. 1.3e-176;
  Matches  507;  Conservative  135;  Mismatches  176;  Indels   81;  Gaps   17;

Qy         24 STPLPQVTSPSFDNAASPVAGRRAVRQTPTSAVRRRGRETDSARRRRSRSRSLGNSVYSS 83
Db         13 NSPPPSVSSPDV-RPSSPLP---ATNSSPPQSGRRGG-----GRRRRGS---------AS 54

Qy         84 PYDAGTPGTPGTPVATPVYATPVGTPMGTPSFHRGTPQYKQRSELGSQGKPLHRRRRSQS 143
Db         55 PYPS-SPSLGG-------FETPPHPGRRTPSGGAAARQQRQ-NWTGGRFPP--------- 96

Qy        144 REPGHRSPSREPSADGRP--SESAEPD--DTLGG---------EYAYVWGTNVNIPDVLR 190
Db         97 ------TPSTPMSTDDVPLSSEAGDEDTPETDGGGGGGAGADATPVFVWGTNISVQDVNA 150

Qy        191 AIRRFLHNYRSSAHDL--------NSKYIQIIEETVERE-EDTLNIDMSDIYDHDPDLYA 241
Db        151 AILRFLRHFR-DPRDAGRVDPVMDEGKYMRAIHRILELEGGESLDVNAHDVFDHDPDLYG 209

Qy        242 KIVRYPLDIIPLLDTECQEVATSLLPTFEKHIEARPFNLKASVHMRELNPSDIDKLVSVK 301
Db        210 KMVRYPLEVLAIFDIVLMDLVARIEPLFEKHIQTRIYNLKSSVCLRNLNPSDIEKMVSIK 269

Qy        302 GMVIRCSSIIPEIKGAFFKCLVCGHSPPLVTVVKGRVEEPTRCEKPECAARNAMSLIHNR 361
Db        270 GMIIRCSSVIPELKEAVFRCLVCGFYSEPVMVDRGRVTEPHICQKEQCKATNSMTLVHNR 329

Qy        362 CTFANKQIVRLQETPDAIPEGETPHTVSMCLYNTMVDAVKPGDRIEVTGVFKAMAVRVGP 421
Db        330 CRFADKQIIKLQETPDEIPEGGTPHTVSVLMHDKLVDAGKPGDRVEITGIYRAMSIRVGP 389

Qy        422 NQRTLRALYKTYIDCVHVKKSDRGRLQTEDPMEMDKENDMYAGYHESDTSEAANEAKIQK 481
Db        390 TQRTVKSIF---------KKTDKSRLHVEDSMETDNPN------ANKTTEDDFLRDKVEK 434

Qy        482 LKELSKLPGIYDRLSRSLAPSIWELEDIKKGLLCQLFGGKAKKIPSGASFRGDINVLLVG 541
```

```
Db        435 LKELSKLPDIYDRLTRSLAPNIWELDDVKRGLLCQLFGGNALRLPSGASFRGDINILLVG 494

Qy        542 DPGTSKSQLLQYVHKIAPRGIYTSGRGSSAVGLTAYVTKDPETRETVLESGALVLSDRGI 601

Db        495 DPGTSKSQLLQYMHKLSPRGIYTSGRGSSAVGLTAYVTKDPETGETVLESGALVLSDKGV 554

Qy        602 CCIDEFDKMSDNARSMLHEVMEQQTVSVAKGGIIASLNARTSVLACANPSGSRYNARLSV 661

Db        555 CCIDEFDKMSDNARSMLHEVMEQQTVSIAKAGIIASLNARTSVLACANPTESRYNPRLSV 614

Qy        662 IDNIQLPPTLLSRFDLIYLMLDKPDEQNDRRLARHLVALHYENYEVSKQDALDLQTLTAY 721

Db        615 IDNIHLPPTLLSRFDLIYLILDKADEQTDRRLAKHIVSLHFENPNIEELEVLDLPTLVAY 674

Qy        722 ITYARQHVHPTLSDEAAEDLINGYVEMRQKGNFPGSSKKVITATPRQLESMIRISEALAR 781

Db        675 ISYARKHIQPQLSDEAAEELTRGYVEMRKRGNSPGSRKKVITATARQIESLIRLSEALAR 734

Qy        782 MRFSEVVEKVDAAEAVRLLDVALQQSATDHATGTIDMDLITTGVSASERIRRANLLAALR 841

Db        735 MRFSEMVEVQDVVEAFRLLEVAMQQSATDHATGTIDMDLIMTGISASERQRRDNLVAATR 794

Qy        842 ELIADKISPGSSSGLKTSQLLEDIRSQSSVDVSLQDIKNALGSLQGEGFLTVHGDIVKR 900

Db        795 NLVMEKMQLGGPS-VRMIELLEEIRKQSSMEVHLHDLRGALGTLMTEGAVVIHGDSVKR 852
```

RESULT 5
US-10-425-115-283561
; Sequence 283561, Application US/10425115
; Publication No. US20040214272A1
; GENERAL INFORMATION:
; APPLICANT: La Rosa, Thomas J.
; APPLICANT: Kovalic, David K.
; APPLICANT: Zhou, Yihua
; APPLICANT: Cao, Yongwei
; TITLE OF INVENTION: Nucleic Acid Molecules and Other Molecules Associated With
; TITLE OF INVENTION: Plants
; FILE REFERENCE: 38-21(53222)B
; CURRENT APPLICATION NUMBER: US/10/425,115
; CURRENT FILING DATE: 2003-04-28
; NUMBER OF SEQ ID NOS: 369326
; SEQ ID NO 283561
; LENGTH: 551
; TYPE: PRT
; ORGANISM: Zea mays
; FEATURE:
; NAME/KEY: unsure
; LOCATION: (1)..(551)
; OTHER INFORMATION: unsure at all Xaa locations
; FEATURE:
; OTHER INFORMATION: Clone ID: MRT4577_21708C.1.pep
US-10-425-115-283561

```
  Query Match             42.8%;  Score 1972.5;  DB 4;  Length 551;
  Best Local Similarity   66.7%;  Pred. No. 3e-139;
  Matches  381;  Conservative   85;  Mismatches   82;  Indels   23;  Gaps    3;

Qy        331 VTVVKGRVEEPTRCEKPECAARNAMSLIHNRCTFANKQIVRLQETPDAIPEGETPHTVSM 390

Db          4 VMVDRGRVTEPHICQKEQCKATNSMTLVHNRCRFSDKQIIKLQETPDEIPEGGTPHTVSV 63

Qy        391 CLYNTMVDAVKPGDRIEVTGVFKAMAVRVGPNQRTLRALYKTYIDCVHVKKSDRGRLQTE 450

Db         64 LMHDKLVDAGKPGDRVEITGIYRAMSIRIGPT----------------XKKTDKSRLHVE 107

Qy        451 DPMEMDKENDMYAGYHESDTSEAANEAKIQKLKELSKLPGIYDRLSRSLAPSIWELEDIK 510

Db        108 DTMDIDNSN------ASKSTEEDFLSDKVEKLKELSKLPDIYERLTRSLAPNIWELDDVK 161

Qy        511 KGLLCQLFGGKAKKIPSGASFRGDINVLLVGDPGTSKSQLLQYVHKIAPRGIYTSGRGSS 570

Db        162 RGLLCQLFGGNPLKLPSGASFRGDINILLVGDPGTSKSQLLQYMHKLSPRGIYTSGRGSS 221

Qy        571 AVGLTAYVTKDPETRETVLESGALVLSDRGICCIDEFDKMSDNARSMLHEVMEQQTVSVA 630

Db        222 AVGLTAYVTKDPETGETVLESGALVLSDKGVCCIDEFDKMSDNARSMLHEVMEQQTVSIA 281

Qy        631 KGGIIASLNARTSVLACANPSGSRYNARLSVIDNIQLPPTLLSRFDLIYLMLDKPDEQND 690

Db        282 KAGIIASLNARTSVLACANPTESRYNPRLSVIDNIHLAPTLLSRFDLIYLILDKADEQTD 341

Qy        691 RRLARHLVALHYENYEVSKQDALDLQTLTAYITYARQHVHPTLSDEAAEDLINGYVEMRQ 750

Db        342 RRLAKHIVSLHFENPNLEELEVLDLQTLVSYISYARKYIQPQLSDEAAEELTRGYVEMRK 401

Qy        751 KGNFPGSSKKVITATPRQLESMIRISEALARMRFSEVVEKVDAAEAVRLLDVALQQSATD 810

Db        402 RGNSPGSRKKVITATARQIESLIRLSEALARMRFSEVVEVRDVVEAFRLLEVAMQQSATD 461

Qy        811 HATGTIDMDLITTGVSASERIRRANLLAALRELIADKISPGSSSGLKTSQLLEDIRSQSS 870

Db        462 HATGTIDMDLIMTGISASERQRRENLVAATRNLIAEKMQLGGPS-MRMIELLEELRKQSS 520

Qy        871 VDVSLQDIKNALGSLQGEGFLTVHGDIVKRV 901

Db        521 MEIHMHELRGALGTLMTEGAVVIHGDNVRRV 551
```

RESULT 6
US-10-425-115-283560
; Sequence 283560, Application US/10425115
; Publication No. US20040214272A1
; GENERAL INFORMATION:
; APPLICANT: La Rosa, Thomas J.
; APPLICANT: Kovalic, David K.
; APPLICANT: Zhou, Yihua
; APPLICANT: Cao, Yongwei
; TITLE OF INVENTION: Nucleic Acid Molecules and Other Molecules Associated With
; TITLE OF INVENTION: Plants
; FILE REFERENCE: 38-21(53222)B
; CURRENT APPLICATION NUMBER: US/10/425,115
; CURRENT FILING DATE: 2003-04-28
; NUMBER OF SEQ ID NOS: 369326
; SEQ ID NO 283560
; LENGTH: 851
; TYPE: PRT
; ORGANISM: Zea mays
; FEATURE:
; NAME/KEY: unsure
; LOCATION: (1)..(851)
; OTHER INFORMATION: unsure at all Xaa locations
; FEATURE:
; OTHER INFORMATION: Clone ID: MRT4577_21707C.1.pep
US-10-425-115-283560

```
  Query Match             41.0%;  Score 1887.5;  DB 4;  Length 851;
  Best Local Similarity   46.2%;  Pred. No. 1.5e-132;
  Matches  423;  Conservative  147;  Mismatches  246;  Indels   99;  Gaps   23;

Qy         20 SEGVSTPLPQVTSPSFDNAASPVAGRRAVRQTPTSAVRRRGRETDSARRRRSRSRSLGNS 79

Db          3 SDGGGSSPPPASSP-YGRPSSPLA---VTNSSPSQPTRRSG-----GRRRRGS------- 46

Qy         80 VYSSPYDAGTPGTPGTPVATPVYATPVGTPMGTPSFHRGTPQYKQRSELGSQGKPLHRRR 139

Db         47 --ASPY-ASSPSLGG-------FETP-------PHPGRRTP-----SGAGA-GAPRQPRQ 83

Qy        140 RSQSREPGHRSPSREPSADG-RPSESAEPDDTLGG-----EYAYVWGTNVNIPDVLRAIR 193

Db         84 NSTGRFP--PTPSTPMSTDDVPPSSEAGDDETDGGGGVDATPVFVWGTNISVQDVNAAIL 141

Qy        194 RFLHNYRSSAHDL--------NSKYIQIIEETVERE-EDTLNIDMSDIYDHDPDLYAKIV 244

Db        142 RFLRHFR-DPRDAGRVDPVMDEGKYMRAIHRILELEGGESLDVDAHDVFDHDPDLYSKMV 200
```

```
Qy        245 RYPLDIIPLLDTECQEVATSLLPTFEKHIEARPFNLKASVHMRELNPSDIDKLVSVKGMV 304
Db        201 RYPLEVLAIFDIVLMDLVARIEPLFEKHIQTRIYNLKSSICLRNLNPSDIEKMVSIKGMI 260

Qy        305 IRCSSIIPEIKGAFFKCLVCGHSPPLVTVVKGRVEEPTRCEKPECAARNAMSLIHNRCTF 364
Db        261 IRCSSVIPELKEAVFRCLVCGFYSEPVMVDRGRVTEPHICQKEQCKATNSMTLVHNRCRF 320

Qy        365 ANKQIVRLQETPDAIPEGETPHTVSMCLYNTMVDAVKPGDRIEVTGVFKAMAVRVGPNQR 424
Db        321 SDKQIIKLQETPDEIPEGGTPHTVSVLMHDKLVDAGKPGDRVEITGIYRAMSIRIGPTQR 380

Qy        425 TLRALYKTYI------DCVHVKKSDRGRLQTEDPMEMDKENDMYAGYHESD---TSEAAN 475
Db        381 TVKSIFKTYIARMRFSEVVEVRDVVEAFRLLEVAMQ-QSATDHATGTIDMDLIMTGISAS 439

Qy        476 E---------AKIQKLKELSKLPGIYDRLSRSLAPSIWELEDIKKGLLCQLFGGKAKKIP 526
Db        440 ERQRRNDLVAATRNLIVEKMQLGG----------PSMHMIEXAVRAQVAELNGEPSARIP 489

Qy        527 SGASFRGDINVLLVGDPGTSKSQLLQYVHKIAPRGIYTSGRGSSAVGLTAYVTKDPETRE 586
Db        490 RGAWHYDDRR--CASDPWKKGEERL----------MLTAICRDGCVSSVYXVSFNSQGCE 537

Qy        587 TVLESGALVLSDRGICCIDEFDKMSDNARSMLHEVMEQQTVSVAKGGIIASLNARTSVLA 646
Db        538 DLIDCRAFAFRVLXVCAXDEFHKMSVNARSMLHEVMEQQTVSIAKAGIIASLNARTSVLA 597

Qy        647 CANPSGSRYNARLSVIDNIQLPPTLLSRFDLIYLMLDKPDEQNDRRLARHLVALHYENYE 706
Db        598 CANPTESRYNPRLSVIDNIHLAPTLLSRFDLIYLILDKADEQTDRRLAKHIVSLHFENPN 657

Qy        707 VSKQDALDLQTLTAYITYARQHVHPTLSDEAAEDLINGYVEMRQKGNFPGSSKKVITATP 766
Db        658 LEELEVLDLQTLVSYISYARKYIQPQLSDEAAEELTRGYVEMRKRGNSPGSRKKVITATA 717

Qy        767 RQLESMIRISEALARMRFSEVVEKVDAAEAVRLLDVALQQSATDHATGTIDMDLITTGVS 826
Db        718 RQIESLIRLSEALARMRFSEVVEVRDVVEAFRLLEVAMQQSATDHATGTIDMDLIMTGIS 777

Qy        827 ASERIRRANLLAALRELIADKISPGSSSGLKTSQLLEDIRSQSSVDVSLQDIKNALGSLQ 886
Db        778 ASERQRRENLVAATRNLIAEKMQLGGPS-MRMIELLEELRKQSSMEIHMHELRGALGTLM 836

Qy        887 GEGFLTVHGDIVKRV 901
Db        837 TEGAVVIHGDNVRRV 851
```

RESULT 7
US-10-128-714-8558
; Sequence 8558, Application US/10128714
; Publication No. US20030119013A1
; GENERAL INFORMATION:
; APPLICANT: Jiang, Bo
; APPLICANT: Hu, Wenqi
; APPLICANT: Tishkoff, Daniel
; APPLICANT: Zamudio, Carlos
; APPLICANT: Eroshkin, Alexey M
; APPLICANT: Lemieux, Sebastien M
; TITLE OF INVENTION: Identification of Essential Genes in Aspergillus fumigatus and
; TITLE OF INVENTION: Methods of Use
; FILE REFERENCE: 10182-018-999
; CURRENT APPLICATION NUMBER: US/10/128,714
; CURRENT FILING DATE: 2002-04-23
; PRIOR APPLICATION NUMBER: US 60/285,697
; PRIOR FILING DATE: 2001-04-23
; PRIOR APPLICATION NUMBER: US 60/287,066
; PRIOR FILING DATE: 2001-04-27
; PRIOR APPLICATION NUMBER: US 60/295,890
; PRIOR FILING DATE: 2001-06-05
; PRIOR APPLICATION NUMBER: US 60/303,899
; PRIOR FILING DATE: 2001-07-09
; PRIOR APPLICATION NUMBER: US 60/316,362
; PRIOR FILING DATE: 2001-08-31
; NUMBER OF SEQ ID NOS: 8603
; SOFTWARE: PatentIn version 3.1
; SEQ ID NO 8558
; LENGTH: 1023
; TYPE: PRT
; ORGANISM: Aspergillus fumigatus
US-10-128-714-8558

```
  Query Match             40.9%;  Score 1883;  DB 4;  Length 1023;
  Best Local Similarity   42.3%;  Pred. No. 4.2e-132;
  Matches  434;  Conservative  155;  Mismatches  286;  Indels  152;  Gaps   23;

Qy         12 VSSPYPSQSEGVSTPLPQVTSPSFDNAASPVAGRRAVRQTPTSAVRRRGRETDSARRRRS 71
Db          1 MSSPASSRRRG---------RPAKDTATSSPARSTRSRQLQTSSPTPRAADEQSQATPRA 51

Qy         72 RSRSLGNSVY--SSPY-----------DAGTPGT----PGTPV----------ATPVYAT 104
Db         52 SRRLRGEAAVPSSSPMFFQSSPSKADSSAETPDVRMDEPSSPMRESSTMDEGDRTPRGNA 111

Qy        105 PV---GTP---MGTPSFHRGTPQYKQRSELGSQGKPLH------------RRRRSQSREP 146
Db        112 PTMRDSSPIRYMSSSSPTRAQNRQSRRSDIPSSSSGLFVSSRPSIESNRAVSRRSDLHSG 171

Qy        147 GH-RSPSRE----------PSADGRPSESA-----EPD----DTLGGEYA-YVWGTNVNI 185
Db        172 GFLSSPNRRRRVFVDANGMPATDGDPRSDATFSNIHPDTSEAEALGGSSTRVIWGTNISI 231

Qy        186 PDVLRAIRRFLHN----YRSSAHDLNSKYIQIIEETVEREE-------------DTLNID 228
Db        232 QDSMSAFKNFLYNFQTKYRLWAEGATEDETRIMGDSAEEREYISMLSTMRQLGVTSLNLD 291

Qy        229 MSDI--YDHDPDLYAKIVRYPLDIIPLLDTECQEVATSL--------------------- 265
Db        292 AKNLKAYPSTLKLWHQLHAYPQEIIPLMDQTVKDVMVELAIKEMERLRAQNQRNQNHNRG 351

Qy        266 -----------------------LPTFEKHIEAR-----PFNLKASVHMRELNPSDIDKL 297
Db        352 LSSGPAVPSSDALSETGRMPQNEIPDLVGEVETKAFKVLPFGLDSTVNMRDLDPADMDKL 411

Qy        298 VSVKGMVIRCSSIIPEIKGAFFKCLVCGHSPPLVTVVKGRVEEPTRCEKPECAARNAMSL 357
Db        412 VSIKGLVIRTTPIIPDMKEAFFRCQVCNHGVQ-VDIDRGKIAEPTECPRPVCKERNSMQL 470

Qy        358 IHNRCTFANKQIVRLQETPDAIPEGETPHTVSMCLYNTMVDAVKPGDRIEVTGVFKAMAV 417
Db        471 IHNRCVFADKQVIKLQETPDSIPDGQTPHSVSLCVYDELVDVCKAGDRVEVTGIFRCNPV 530

Qy        418 RVGPNQRTLRALYKTYIDCVHVKKSDRGRLQTEDPMEMDKENDMYAGYHESDTSEAANEA 477
Db        531 RVNPRQRTQKSLFKTYIDVLHVQKIDRKKLGIDVSTIEQELSEQAAGDAEQTRRLTAEEE 590

Qy        478 KIQKLKELSKLPGIYDRLSRSLAPSIWELEDIKKGLLCQLFGGKAKKIPSGAS--FRGDI 535
Db        591 --EKIKRTATRPDLYELLSRSLAPSIYEMDDVKKGILLQLFGGTNKTFQKGGNPRYRGDI 648

Qy        536 NVLLVGDPGTSKSQLLQYVHKIAPRGIYTSGRGSSAVGLTAYVTKDPETRETVLESGALV 595
Db        649 NILLCGDPSTSKSQLLRYVHKIAPRGVYTSGKGSSAVGLTAYVTRDPETRQMVLESGALV 708

Qy        596 LSDRGICCIDEFDKMSDNARSMLHEVMEQQTVSVAKGGIIASLNARTSVLACANPSGSRY 655
Db        709 LSDGGICCIDEFDKMNESTRSVLHEVMEQQTVSIAKAGIITTLNARTSILASANPIGSRY 768

Qy        656 NARLSVIDNIQLPPTLLSRFDLIYLMLDKPDEQNDRRLARHLVALHYEN--YEVSKQDAL 713
Db        769 NPNLPVPQNIDLPPTLLSRFDLVYLVLDRVDEQEDRRLAKHLVNMYLEDRPEHAAEQEIL 828

Qy        714 DLQTLTAYITYARQHVHPTLSDEAAEDLINGYVEMRQKGNFPGSSKKVITATPRQLESMI 773
```

```
Db        829 PIEFLTAYITYAKTKVHPVLTPAAGKALSDAYVNMRKLGDDIRSSDRRITATTRQLESMI 888

Qy        774 RISEALARMRFSEVVEKVDAAEAVRLLDVALQQSATDHATGTIDMDLITTGVSASERIRR 833
              |:||| |||| |   |    |  |||||:  |::|:|||  || ||| |:| | |||||  |
Db        889 RLSEAHARMRLSPEVTADDVEEAVRLIRSAIKQAATDSRTGLIDMSLLTEGTSASERRSR 948

Qy        834 ANLLAALRELIADKISPGSSSGLKTSQLLEDIRSQSSVDVSLQDIKNALGSLQGEGFLTV 893
                |  ||  :: |  | | ::  : :::   :    ||::|        :|: :|: || ::|
Db        949 EALKRALLSVVDDLCSGGGAA--RWAEVFRILSENSSIEVDGAQFADAVRALEAEGAVSV 1006

Qy        894 HGDIVKR 900
               |:  :|
Db       1007 VGEGARR 1013
```

RESULT 8
US-10-128-714-3558
; Sequence 3558, Application US/10128714
; Publication No. US20030119013A1
; GENERAL INFORMATION:
; APPLICANT: Jiang, Bo
; APPLICANT: Hu, Wenqi
; APPLICANT: Tishkoff, Daniel
; APPLICANT: Zamudio, Carlos
; APPLICANT: Eroshkin, Alexey M
; APPLICANT: Lemieux, Sebastien M
; TITLE OF INVENTION: Identification of Essential Genes in Aspergillus fumigatus and
; TITLE OF INVENTION: Methods of Use
; FILE REFERENCE: 10182-018-999
; CURRENT APPLICATION NUMBER: US/10/128,714
; CURRENT FILING DATE: 2002-04-23
; PRIOR APPLICATION NUMBER: US 60/285,697
; PRIOR FILING DATE: 2001-04-23
; PRIOR APPLICATION NUMBER: US 60/287,066
; PRIOR FILING DATE: 2001-04-27
; PRIOR APPLICATION NUMBER: US 60/295,890
; PRIOR FILING DATE: 2001-06-05
; PRIOR APPLICATION NUMBER: US 60/303,899
; PRIOR FILING DATE: 2001-07-09
; PRIOR APPLICATION NUMBER: US 60/316,362
; PRIOR FILING DATE: 2001-08-31
; NUMBER OF SEQ ID NOS: 8603
; SOFTWARE: PatentIn version 3.1
; SEQ ID NO 3558
;   LENGTH: 874
;   TYPE: PRT
;   ORGANISM: Aspergillus fumigatus
US-10-128-714-3558

```
  Query Match             40.4%;  Score 1860;  DB 4;  Length 874;
  Best Local Similarity   45.5%;  Pred. No. 1.8e-130;
  Matches  413;  Conservative  146;  Mismatches  244;  Indels  104;  Gaps   18;

Qy         51 TPTSAVRRRGRETDSARRRRSRSRSLGNSVYSSPYDAGTPGTPGTPVATPVYATPVGTPM 110
              :||  |  |: | :|          | | |   | | |
Db          5 SPTRAQNRQSRRSDIP------SSSSGLFVSSRP------------------------- 32

Qy        111 GTPSFHRGTPQYKQRSELGSQG--KPLHRRRRSQSREPGHRSPSREPSADGRPSESA--- 165
                :   :|        :||:| | |      :||||      |          |: || |   |
Db         33 -SIESNRAV---SRRSDLHSGGFLSSPNRRRRVFVDANG------MPATDGDPRSDATFS 82

Qy        166 --EPD----DTLGGEYA-YVWGTNVNIPDVLRAIRRFLHN----YRSSAHDLNSKYIQII 214
                 ||      : |||       :||||::|  | :  | :  ||:|      ||  |          :|:
Db         83 NIHPDTSEAEALGGSSTRVIWGTNISIQDSMSAFKNFLYNFQTKYRLWAEGATEDETRIM 142

Qy        215 EETVEREE-------------DTLNIDMSDI--YDHDPDLYAKIVRYPLDIIPLLDTECQ 259
               :: |  |                 :||:|   ::  |         |: ::   || :||||:|    :
Db        143 GDSAEEREYISMLSTMRQLGVTSLNLDAKNLKAYPSTLKLWHQLHAYPQEIIPLMDQTVK 202

Qy        260 EVATSL---------------------LPTFE-KHIEARPFNLKASVHMRELNPSDIDKL 297
              :|    |                     | :  | |   |  :   || |  ::|:|:||:|:|:|:|||
Db        203 DVMVELAIKEMERLRAQNQRNQNHNRGLSSVETKAFKVLPFGLDSTVNMRDLDPADMDKL 262

Qy        298 VSVKGMVIRCSSIIPEIKGAFFKCLVCGHSPPLVTVVKGRVEEPTRCEKPECAARNAMSL 357
              ||:||:|||  :  |||::|  |||:|  ||  |        |  : :|::  |||  |  :|  |     ||:|  |
Db        263 VSIKGLVIRTTPIIPDMKEAFFRCQVCNHGVQ-VDIDRGKIAEPTECPRPVCKERNSMQL 321

Qy        358 IHNRCTFANKQIVRLQETPDAIPEGETPHTVSMCLYNTMVDAVKPGDRIEVTGVFKAMAV 417
              |||||  ||:||:::||||||:||:|:||:|:||:||:|:   :||    | |||:|||:|:    |
Db        322 IHNRCVFADKQVIKLQETPDSIPDGQTPHSVSLCVYDELVDVCKAGDRVEVTGIFRCNPV 381

Qy        418 RVGPNQRTLRALYKTYIDCVHVKKSDRGRLQTEDPMEMDKENDMYAGYHESDTSEAANEA 477
              || |  |||  ::|:||||  :||:|  ||  :|     :       : ::     ||     |        | |
Db        382 RVNPRQRTQKSLFKTYIDVLHVQKIDRKKLGIDVSTIEQELSEQAAGDAEQTRRLTAEEE 441

Qy        478 KIQKLKELSKLPGIYDRLSRSLAPSIWELEDIKKGLLCQLFGGKAKKIPSGAS--FRGDI 535
                :|:|   :    |  :|:  ||||||||||:|::|:|||:|  |||||   |         |  :   :||||
Db        442 --EKIKRTATRPDLYELLSRSLAPSIYEMDDVKKGILLQLFGGTNKTFQKGGNPRYRGDI 499

Qy        536 NVLLVGDPGTSKSQLLQYVHKIAPRGIYTSGRGSSAVGLTAYVTKDPETRETVLESGALV 595
              |:||  |||  ||||||||:||||||||:|||||:|||||||||||||:|||||:  ||||||||
Db        500 NILLCGDPSTSKSQLLRYVHKIAPRGVYTSGKGSSAVGLTAYVTRDPETRQMVLESGALV 559

Qy        596 LSDRGICCIDEFDKMSDNARSMLHEVMEQQTVSVAKGGIIASLNARTSVLACANPSGSRY 655
              |||  |||||||||||:::  ||:||||||||||||:|| |||  ||   :||||||:||  |||  ||||
Db        560 LSDGGICCIDEFDKMNESTRSVLHEVMEQQTVSIAKAGIITTLNARTSILASANPIGSRY 619

Qy        656 NARLSVIDNIQLPPTLLSRFDLIYLMLDKPDEQNDRRLARHLVALHYEN--YEVSKQDAL 713
              |    |   ||  |||||||||||:||:||:  |||  ||||:|||  :: |:           ::|:  |
Db        620 NPNLPVPQNIDLPPTLLSRFDLVYLVLDRVDEQEDRRLAKHLVNMYLEDRPEHAAEQEIL 679

Qy        714 DLQTLTAYITYARQHVHPTLSDEAAEDLINGYVEMRQKGNFPGSSKKVITATPRQLESMI 773
               :: ||||||||:   |||  |:   |   |  :  ||  ||:  |:        ||  :  |||| ||||||
Db        680 PIEFLTAYITYAKTKVHPVLTPAAGKALSDAYVNMRKLGDDIRSSDRRITATTRQLESMI 739

Qy        774 RISEALARMRFSEVVEKVDAAEAVRLLDVALQQSATDHATGTIDMDLITTGVSASERIRR 833
              |:||| |||| |   |    |  |||||:  |::|:|||  || ||| |:| | |||||  |
Db        740 RLSEAHARMRLSPEVTADDVEEAVRLIRSAIKQAATDSRTGLIDMSLLTEGTSASERRSR 799

Qy        834 ANLLAALRELIADKISPGSSSGLKTSQLLEDIRSQSSVDVSLQDIKNALGSLQGEGFLTV 893
                |  ||  :: |  | | ::  : :::   :    ||::|        :|: :|: || ::|
Db        800 EALKRALLSVVDDLCSGGGAA--RWAEVFRILSENSSIEVDGAQFADAVRALEAEGAVSV 857

Qy        894 HGDIVKR 900
               |:  :|
Db        858 VGEGARR 864
```

RESULT 9
US-10-320-797-3332
; Sequence 3332, Application US/10320797
; Publication No. US20040014955A1
; GENERAL INFORMATION:
; APPLICANT: Eroshkin, Alexey M.
; APPLICANT: Zamudio, Carlos
; TITLE OF INVENTION: IDENTIFICATION OF ESSENTIAL GENES OF CRYPTOCOCCUS NEOFORMANS AND
; TITLE OF INVENTION: METHODS OF USE
; FILE REFERENCE: 10182-021-999
; CURRENT APPLICATION NUMBER: US/10/320,797
; CURRENT FILING DATE: 2002-12-16
; PRIOR APPLICATION NUMBER: 60/341,261
; PRIOR FILING DATE: 2001-12-17
; NUMBER OF SEQ ID NOS: 3361
; SOFTWARE: PatentIn version 3.1
; SEQ ID NO 3332
;   LENGTH: 989
;   TYPE: PRT
;   ORGANISM: Cryptococcus neoformans
US-10-320-797-3332

```
  Query Match             39.8%;  Score 1835.5;  DB 4;  Length 989;
  Best Local Similarity   43.8%;  Pred. No. 1.5e-128;
  Matches  409;  Conservative  147;  Mismatches  246;  Indels  131;  Gaps   20;

Qy         47 AVRQTPTSAVRR-----RGRETDSARRRRSRSRSLGNSVYSSP---------------YD 86
              |: |:  ||::       ||     |   ||    |       :  :|||
```

```
Db          4 ALLQSDESAMKHHCSSLRGSSGGSTPRRARRGDIHSSFPHSSPSLARRQAQVPADSLSLR 63

Qy         87 AGTPGTPGTPVATPVYATPVGTPMGTPSFHRGTPQYKQRSELGSQGKPLHRRRRSQSREP 146
Db         64 ASSPG-----------------MGTDSF--GTPRV-----FGSAAPTLSAVAQSQL--- 95

Qy        147 GHRSPSREPSADGRPSESAEPDDTLGGEYAYVWGTNVNIPDVLRAIRRFL---------- 196
Db         96 ---------GADG---------DDIDGMVKFIWGTTISLQESMNLFRDFLRGFKPKYRAV 137

Qy        197 HNYRSSAHDLNS---------KYIQIIEETVE------------REEDTLNIDMSDIYD 234
Db        138 YNAEQSRHAEESGGVAPPPMTLYDNLSAERAEVPLYETYLNRLRLTGETNLNLDALNLLA 197

Qy        235 HDP--DLYAKIVRYPLDIIPLLDTECQEVAT-----------------SLLPTFEKHI-E 274
Db        198 YRPTKKLYQQLVNYPQEVIPIMDQVLRDVMIELGHEELEKAQTKFAEGNLSQDVESRVYK 257

Qy        275 ARPFNLKASVHMRELNPSDIDKLVSVKGMVIRCSSIIPEIKGAFFKCLVCGHSPPLVTVV 334
Db        258 VRPFGGEKTVNMRDLNPGDTDKLVTVKGLVIRATPVIPDMTTAFFRCLVCQHTVQ-ADID 316

Qy        335 KGRVEEPTRCEKPECAARNAMSLIHNRCTFANKQIVRLQETPDAIPEGETPHTVSMCLYN 394
Db        317 RGRISEPERCPRDVCGSTGTMSLIHNRSEFTDKQVIRLQETPDAVPDGQTPHTVSLCVYD 376

Qy        395 TMVDAVKPGDRIEVTGVFKAMAVRVGPNQRTLRALYKTYIDCVHVKKSDRGRLQTEDPME 454
Db        377 ELVDLVKPGDRVIITGIFRSIPVRVNPRQRSIKSLYKTYLDVVHVKRTNTARMGF-DPST 435

Qy        455 MDKENDMYAGYHESDTSEAANEAKIQKLKELSKLPGIYDRLSRSLAPSIWELEDIKKGLL 514
Db        436 RAGESKP-PGDLGSPVRSAAAEME-QRIIELSNHPDLYNILASSLAPSIYELEDVKKGIL 493

Qy        515 CQLFGGKAKKIPSGAS-----FRGDINVLLVGDPGTSKSQLLQYVHKIAPRGIYTSGRGS 569
Db        494 LQLFGGTNKSIARGGGGGGPRYRGDINVLMVGDPGTSKSQILQYVHKIAPRGVYTSGKGS 553

Qy        570 SAVGLTAYVTKDPETRETVLESGALVLSDRGICCIDEFDKMSDNARSMLHEVMEQQTVSV 629
Db        554 SAVGLTAYVTRDPDSKQLVLESGALVLSDGGVCCIDEFDKMSDATRSVLHEVMEQQTVSI 613

Qy        630 AKGGIIASLNARTSVLACANPSGSRYNARLSVIDNIQLPPTLLSRFDLIYLMLDKPDEQN 689
Db        614 AKAGIITTLNARTSILAAANPINSRYDPNLPIPANIDLPPTLISRFDLLYLVLDKVDEVN 673

Qy        690 DRRLARHLVALHYENYEVSKQD-ALDLQTLTAYITYARQHVHPTLSDEAAEDLINGYVEM 748
Db        674 DRKLAKHLVGLYLSDVEDQPADNIIPLQTLTSYITYARSKIHPVLTEGASEALVQAYVEM 733

Qy        749 RQKGNFPGSSKKVITATPRQLESMIRISEALARMRFSEVVEKVDAAEAVRLLDVALQQSA 808
Db        734 RKAGMDSRTQEKRITATTRQLESMIRLGEAHARMRLSDRVEEEDIREAVRLIKSALRESA 793

Qy        809 TDHATGTIDMDLITTGVSASERIRRANLLAALRELIADKISPGSSSGLKTSQLLEDIRSQ 868
Db        794 TDPLTGQIDLDLINTGAGQTMRRARADLKREVIKLVVEK---ARSQGIRWAAVIDELNKQ 850

Qy        869 SSVDVSLQDIKNALGSLQGEGFLTVHGDIVKRV 901
Db        851 SSVPVDHAQFAEIVRELEEESIVKVMGERERRI 883
```

RESULT 10
US-10-369-493-3801
; Sequence 3801, Application US/10369493
; Publication No. US20030233675A1
; GENERAL INFORMATION:
; APPLICANT: Cao, Yongwei
; APPLICANT: Hinkle, Gregory J.
; APPLICANT: Slater, Steven C.
; APPLICANT: Goldman, Barry S.
; APPLICANT: Chen, Xianfeng
; TITLE OF INVENTION: EXPRESSION OF MICROBIAL PROTEINS IN PLANTS FOR PRODUCTION OF
; TITLE OF INVENTION: PLANTS WITH IMPROVED PROPERTIES
; FILE REFERENCE: 38-10(52052)B
; CURRENT APPLICATION NUMBER: US/10/369,493
; CURRENT FILING DATE: 2003-02-28
; PRIOR APPLICATION NUMBER: US 60/360,039
; PRIOR FILING DATE: 2002-02-21
; NUMBER OF SEQ ID NOS: 47374
; SEQ ID NO 3801
; LENGTH: 960
; TYPE: PRT
; ORGANISM: Neurospora crassa
; FEATURE:
; NAME/KEY: unsure
; LOCATION: (1)..(960)
; OTHER INFORMATION: unsure at all Xaa locations
US-10-369-493-3801

Query Match 39.8%; Score 1835; DB 4; Length 960;
Best Local Similarity 41.2%; Pred. No. 1.6e-128;
Matches 412; Conservative 163; Mismatches 248; Indels 178; Gaps 23;

```
Qy         24 STPLPQVTSPSFD-------NAASPVAGRRAVRQTPTSAVRRRGRETDSARRRRSRSRSL 76
Db          1 SSPLRQQTETQSDADRTPRANGRSQLIGGMAVNQ--------------KKKKKDYTECX 45

Qy         77 GNSVYSSPYDAGTPGTPGTPVATPVYATPVGTPMGTPSFHRGTPQYKQRSE-----LGSQ 131
Db         46 PHSADSSPIRYASSSSPGRQL--------------------TQQSDLRSESSQLFVSSQ 84

Qy        132 GKPLHRRRRSQSREPGHRSPSR----------------EPSADGRPSESAEPD----DTL 171
Db         85 RSVAGRSRRGDINGDPLRTPAQIPRRIILDDAGRVIRDAPGSDANSFVTNNPNTSEADAL 144

Qy        172 GGE-YAYVWGTNVNIPDVLRAIRRFLHN----YRSSAH-----------DLNSK-YIQII 214
Db        145 GGQSQGLVWGTTISLDDSFSAFKDFLRNFTRKYRMWADGADEAETIGHPDADSKPYWEAL 204

Qy        215 EETVEREEDTLNIDMSDI--YDHDPDLYAKIVRYPLDIIPLLDT---EC------QEVAT 263
Db        205 ENMLLLGTNKLYLDLRDLKSYPRTLKLWHQAQHYPTEIIPVMDQCVHDCMMELAQKEMAS 264

Qy        264 -------------------------------------SLLPTFEKHIE-----ARPFNLK 281
Db        265 QRASQNSRTAPGASQSSEPNFPSSERSEEPPTPRPAQTAAPTIEDQVSQMAYVVRPWGLD 324

Qy        282 ASVHMRELNPSDIDKLVSVKGMVIRCSSIIPEIKGAFFKCLVCGHSPPLVTVV--KGRVE 339
Db        325 KITNLRDLNPSDMDKLVSIKGLVIRTTPVIPDMKDAFFKCSVCGHS---ITVQLDRGKIR 381

Qy        340 EPTRCEKPECAARNAMSLIHNRCTFANKQIVRLQETPDAIPEGETPHTVSMCLYNTMVDA 399
Db        382 EPTECPRARCASKNSMQIIHNRCAFEDKQVIKLQETPDNVPAGQTPHSVSVCVYNELVDF 441

Qy        400 VKPGDRIEVTGVFKAMAVRVGPNQRTLRALYKTYIDCVHVKKSDRGRLQTEDPMEMD--K 457
Db        442 CKAGDRVELTGIFKVTPVRVNPRMRTVKSVHKTYVDVVHVQKVDRKRMGS-DPSTLDLAE 500

Qy        458 ENDMYAGYHESDTSEAANEAKIQKLKELSKLPGIYDRLSRSLAPSIWELEDIKKGLLCQL 517
Db        501 EEEAHANGQSMDEVRKVSPDEEERIKETAARPDIYDLLSRSLAPSIYEMDDVKKGILLQL 560

Qy        518 FGGKAKKIPSGAS--FRGDINVLLVGDPGTSKSQLLQYVHKIAPRGIYTSGRGSSAVGLT 575
Db        561 FGGTNKTFEKGGSPKYRGDINVLLCGDPSTSKSQLLSYVHRIAPRGVYTSGKGSSAVGLT 620

Qy        576 AYVTKDPETRETVLESGALVLSDRGICCIDEFDKMSDNARSMLHEVMEQQTVSVAKGGII 635
Db        621 AYVTRDPESRQLVLESGALVLSDGGVCCIDEFDKMNESTRSVLHEVMEQQTVSVAKAGII 680

Qy        636 ASLNARTSVLACANPSGSRYNARLSVIDNIQLPPTLLSRFDLIYLMLDKPDEQNDRRLAR 695
Db        681 TTLNARTSILASANPIGSRYNPDLSVPQNIDLPPTLLSRFDLVYLILDRVDEKNDQRLAR 740
```

```
Qy        696 HLVALHYENYEVSKQDALDL--------------------------QTLTAYITYARQH 728
Db        741 HLLSMYLEDKPESAQQANDVLVSFSPSVNPHEDEPQTNQLPLKKTARRVLTSYISYARSH 800

Qy        729 VHPTLSDEAAEDLINGYVEMRQKGNFPGSSKKVITATPRQLESMIRISEALARMRFSEVV 788
Db        801 IHPALTPEAGRELVDAYVEMRKLGQDVRAAEKRITATTRQLESMIRLAEAHAKMRLSQTV 860

Qy        789 EKVDAAEAVRLLDVALQQSATDHATGTIDMDLITTGVSASERIRRANLLAALRELIADKI 848
Db        861 TRDDVREAVRLIKSALKTAATD-SQGRIDMSLLTEGTSAAERQRKADMKDAVIRLLDEMT 919

Qy        849 SPGSSSGLKTSQLLEDIRSQSSVDVSLQDIKNALGSLQGEG 889
Db        920 SGGQV--VRYSEVARRLGEGAGVQVEPAEFAEVMRALEMEG 958
```

RESULT 11
US-10-369-493-22511
; Sequence 22511, Application US/10369493
; Publication No. US20030233675A1
; GENERAL INFORMATION:
; APPLICANT: Cao, Yongwei
; APPLICANT: Hinkle, Gregory J.
; APPLICANT: Slater, Steven C.
; APPLICANT: Goldman, Barry S.
; APPLICANT: Chen, Xianfeng
; TITLE OF INVENTION: EXPRESSION OF MICROBIAL PROTEINS IN PLANTS FOR PRODUCTION OF
; TITLE OF INVENTION: PLANTS WITH IMPROVED PROPERTIES
; FILE REFERENCE: 38-10(52052)B
; CURRENT APPLICATION NUMBER: US/10/369,493
; CURRENT FILING DATE: 2003-02-28
; PRIOR APPLICATION NUMBER: US 60/360,039
; PRIOR FILING DATE: 2002-02-21
; NUMBER OF SEQ ID NOS: 47374
; SEQ ID NO 22511
;   LENGTH: 933
;   TYPE: PRT
;   ORGANISM: Saccharomyces cerevisiae
US-10-369-493-22511

```
  Query Match             39.8%;  Score 1833;  DB 4;  Length 933;
  Best Local Similarity   40.9%;  Pred. No. 2.1e-128;
  Matches  393;  Conservative  178;  Mismatches  281;  Indels  110;  Gaps   14;

Qy          2 ENNDALDIGAVSSPYPSQSEGVSTPLPQVTSPSFDNAASPVAG---------------- 44
Db         11 EDNNS------SSPVVPNPDSVP---PQLSSPALFYSSSSSQGDIYGRNNSQNLSQGEGN 61

Qy         45 -RRAVRQTP---TSAVRRRGRETDSARRRRSRSRSLGNSVYSSPYDAGTPGTPGTPVATP 100
Db         62 IRAAIGSSPLNFPSSSQRQNSDVFQSQGRQGRIRSSASASGRSRYHSDLRSDRALPTSS- 120

Qy        101 VYATPVGTPMGTPSFHRGTPQYKQRSELGSQGKPLHRRRRSQSREPGHRSPSREPSADGR 160
Db        121 -------------------------SSLGRNGQNRVHMRRNDIHTSDLSSPRRIVDFDTR 155

Qy        161 --------------PSESAEPDDTLGGEYAYVWGTNVNIPDVLRAIRRFLHNYR------ 200
Db        156 SGVNTLDTSSSSAPPSEASEP-------LRIIWGTNVSIQECTTNFRNFLMSFKYKFRKI 208

Qy        201 --------SSAHDLNSKYIQIIEETVEREEDTLNIDMSDI--YDHDPDLYAKIVRYPLDI 250
Db        209 LDEREEFINNTTDEELYYIKQLNEMRELGTSNLNLDARNLLAYKQTEDLYHQLLNYPQEV 268

Qy        251 IPLLDTECQEVATSL---------LPTFE-KHIEARPFNLKASVHMRELNPSDIDKLVSV 300
Db        269 ISIMDQTIKDCMVSLIVDNNLDYDLDEIETKFYKVRPYNVGSCKGMRELNPNDIDKLINL 328

Qy        301 KGMVIRCSSIIPEIKGAFFKCLVCGHSPPLVTVVKGRVEEPTRCEKPECAARNAMSLIHN 360
Db        329 KGLVLRSTPVIPDMKVAFFKCNVCDHT-MAVEIDRGVIQEPARCERIDCNEPNSMSLIHN 387

Qy        361 RCTFANKQIVRLQETPDAIPEGETPHTVSMCLYNTMVDAVKPGDRIEVTGVFKAMAVRVG 420
Db        388 RCSFADKQVIKLQETPDFVPDGQTPHSISLCVYDELVDSCRAGDRIEVTGTFRSIPIRAN 447

Qy        421 PNQRTLRALYKTYIDCVHVKKSDRGRLQTEDPMEMDKENDMYAGYHESDTSEAANEAKIQ 480
Db        448 SRQRVLKSLYKTYVDVVHVKKVSDKRLDVDTSTIEQELMQNKVDHNEVEEVRQITDQDLA 507

Qy        481 KLKELSKLPGIYDRLSRSLAPSIWELEDIKKGLLCQLFGGKAKKIPSGASFRGDINVLLV 540
Db        508 KIREVAAREDLYSLLARSIAPSIYELEDVKKGILLQLFGGTNKTFTKGGRYRGDINILLC 567

Qy        541 GDPGTSKSQLLQYVHKIAPRGIYTSGRGSSAVGLTAYVTKDPETRETVLESGALVLSDRG 600
Db        568 GDPSTSKSQILQYVHKITPRGVYTSGKGSSAVGLTAYITRDVDTKQLVLESGALVLSDGG 627

Qy        601 ICCIDEFDKMSDNARSMLHEVMEQQTVSVAKGGIIASLNARTSVLACANPSGSRYNARLS 660
Db        628 VCCIDEFDKMSDSTRSVLHEVMEQQTISIAKAGIITTLNARSSILASANPIGSRYNPNLP 687

Qy        661 VIDNIQLPPTLLSRFDLIYLMLDKPDEQNDRRLARHLVALHYEN--YEVSKQDALDLQTL 718
Db        688 VTENIDLPPPLLSRFDLVYLVLDKVDEKNDRELAKHLTNLYLEDKPEHISQDDVLPVEFL 747

Qy        719 TAYITYARQHVHPTLSDEAAEDLINGYVEMRQKGNFPGSSKKVITATPRQLESMIRISEA 778
Db        748 TMYISYAKEHIHPIITEAAKTELVRAYVGMRKMGDDSRSDEKRITATTRQLESMIRLAEA 807

Qy        779 LARMRFSEVVEKVDAAEAVRLLDVALQQSATDHATGTIDMDLITTGVSASERIRRANLLA 838
Db        808 HAKMKLKNVVELEDVQEAVRLIRSAIKDYATDPKTGKIDMNLVQTGKSVIQRKLQEDLSR 867

Qy        839 ALRELIADKISPGSSSGLKTSQLLEDIRSQSSVDVSLQDIKNALGSLQGEGFLTVHGDIV 898
Db        868 EIMNVLKDQ----ASDSMSFNELIKQINEHSQDRVESSDIQEALSRLQQEDKVIVLGEGV 923

Qy        899 KR 900
Db        924 RR 925
```

RESULT 12
US-10-756-149-5246
; Sequence 5246, Application US/10756149
; Publication No. US20050181375A1
; GENERAL INFORMATION:
; APPLICANT: Aziz, Natasha
; APPLICANT: Zlotnik, Albert
; TITLE OF INVENTION: NOVEL METHODS OF DIAGNOSIS OF METASTATIC CANCER, COMPOSITIONS AND
; TITLE OF INVENTION: METHODS OF SCREENING FOR MODULATORS OF METASTATIC CANCER
; FILE REFERENCE: file
; CURRENT APPLICATION NUMBER: US/10/756,149
; CURRENT FILING DATE: 2004-01-12
; NUMBER OF SEQ ID NOS: 5818
; SOFTWARE: PatentIn version 3.2
; SEQ ID NO 5246
;   LENGTH: 923
;   TYPE: PRT
;   ORGANISM: Homo Sapiens
US-10-756-149-5246

```
  Query Match             39.4%;  Score 1817.5;  DB 5;  Length 923;
  Best Local Similarity   43.7%;  Pred. No. 3e-127;
  Matches  404;  Conservative  159;  Mismatches  275;  Indels   86;  Gaps   19;

Qy         18 SQSEGVSTPLP-QVTSPSFDNAASPVA-----GRRAVRQTPTSAVRRR-GRETDSARRRR 70
Db         42 SREAGPARACPCRAGTPS--TMSSPASTPSRRGSRRGRATPAQTPRSEDARSSPSQRRRG 99

Qy         71 SRSRSLG-------------------NSVYSSPYDAGTPGTP-GTPVATPV-YATPVGTP 109
Db        100 EDSTSTGELQPMPTSPGVDLQSTAAQDVLFSSPPQMHSSAIPLDFDVSSPLTYGTPSSRV 159
```

```
Qy    110 MGTP-SFHRGTPQYKQRSELGSQGKPLHRRRRSQSREPGHRSPSREPSADGRPSESAEPD 168
Db    160 EGTPRSGVRGTP-VRQRPDLGSAQKGLQVDLQS---------------DGAAAEDIVAS 202

Qy    169 DTLGGEYAYVWGTNVNIPDVLRAIRRFLHNY--------RSSAHDLNSK-YIQIIEETVE 219
Db    203 EQSLGQKLVIWGTDVNVAACKENFQRFLQRFIDPLAKEEENVGIDITEPLYMQRLGEINV 262

Qy    220 REEDTLNIDMSDIYDHDPDLYAKIVRYPLDIIPLLDTECQEVATSLLP--TFEKHIEARP 277
Db    263 IGEQFLNVNCEHIKSFDKNLYRQLISYPQEVIPTFDMAVNEIFFDRYPDSILEHQIQVRP 322

Qy    278 FNLKASVHMRELNPSDIDKLVSVKGMVIRCSSIIPEIKGAFFKCLVCGHSPPLVTVVKGR 337
Db    323 FNALKTKNMRNLNPEDIDQLITISGMVIRTSQLIPEMQEAFFQCQVCAHT-TRVEMDRGR 381

Qy    338 VEEPTRCEKPECAARNAMSLIHNRCTFANKQIVRLQETPDAIPEGETPHTVSMCLYNTMV 397
Db    382 IAEPSVCGR--CHTTHSMALIHNRSLFSDKQMIKLQESPEDMPAGQTPHTVILFAHNDLV 439

Qy    398 DAVKPGDRIEVTGVFKAMAVRVGPNQRTLRALYKTYIDCVHVKKSDRGRLQTEDPMEMDK 457
Db    440 DKVQPGDRVNVTGIYRAVPIRVNPRVSNVKSVYKTHIDVIHYRKTDAKRLH--------- 490

Qy    458 ENDMYAGYHESDTSEAANEAKIQKLKELSKLPGIYDRLSRSLAPSIWELEDIKKGLLCQL 517
Db    491 ------GLDEEAEQKLFSEKRVELLKELSRKPDIYERLASALAPSIYEHEDIKKGILLQL 544

Qy    518 FGGKAKKI--PSGASFRGDINVLLVGDPGTSKSQLLQYVHKIAPRGIYTSGRGSSAVGLT 575
Db    545 FGGTRKDFSHTGRGKFRAEINILLCGDPGTSKSQLLQYVYNLVPRGQYTSGKGSSAVGLT 604

Qy    576 AYVTKDPETRETVLESGALVLSDRGICCIDEFDKMSDNARSMLHEVMEQQTVSVAKGGII 635
Db    605 AYVMKDPETRQLVLQTGALVLSDNGICCIDEFDKMNESTRSVLHEVMEQQTLSIAKAGII 664

Qy    636 ASLNARTSVLACANPSGSRYNARLSVIDNIQLPPTLLSRFDLIYLMLDKPDEQNDRRLAR 695
Db    665 CQLNARTSVLAAANPIESQWNPKKTTIENIQLPHTLLSRFDLIFLMLDPQDEAYDRRLAH 724

Qy    696 HLVALHYENYEVSKQDALDLQTLTAYITYARQHVHPTLSDEAAEDLINGYVEMRQKGNFP 755
Db    725 HLVALYYQSEEQAEEELLDMAVLKDYIAYAHSTIMPRLSEEASQALIEAYVDMRK----I 780

Qy    756 GSSKKVITATPRQLESMIRISEALARMRFSEVVEKVDAAEAVRLLDVALQQSATDHATGT 815
Db    781 GSSRGMVSAYPRQLESLIRLAEAHAKVRLSNKVEAIDVEEAKRLHREALKQSATDPRTGI 840

Qy    816 IDMDLITTGVSASERIRRANLLAALRELIADKISPGSSSGLKTSQLLEDIRSQSSVDVSL 875
Db    841 VDISILTTGMSATSRKRKEELAEALKKLI---LSKGKTPALKYQQLFEDIRGQSDIAITK 897

Qy    876 QDIKNALGSLQGEGFLTVHGDIVK 899
Db    898 DMFEEALRALADDDFLTVTGKTVR 921
```

RESULT 13
US-10-032-585-7480
; Sequence 7480, Application US/10032585
; Publication No. US20030180953A1
; GENERAL INFORMATION:
; APPLICANT: Terry, Roemer D.
; APPLICANT: Bo, Jiang
; APPLICANT: Charles, Boone
; APPLICANT: Howard, Bussey
; TITLE OF INVENTION: Gene Disruption Methodologies for Drug Target Discovery
; FILE REFERENCE: 10182-005-999
; CURRENT APPLICATION NUMBER: US/10/032,585
; CURRENT FILING DATE: 2001-12-20
; NUMBER OF SEQ ID NOS: 8000
; SOFTWARE: PatentIn version 3.1
; SEQ ID NO 7480
; LENGTH: 910
; TYPE: PRT
; ORGANISM: Candida albicans
US-10-032-585-7480

Query Match 38.1%; Score 1754; DB 4; Length 910;
Best Local Similarity 42.0%; Pred. No. 1.8e-122;
Matches 397; Conservative 155; Mismatches 259; Indels 134; Gaps 20;

```
Qy      3 NNDALDIGAVSSPYPSQSE---GVSTPLPQVTSPSFDNAASPVAGRRAVRQTPTSAV--- 56
Db     45 SNPGSDIGNFNSQSQSQSQRRNDISSPLHYTSSAQ----------------PTSDIGGF 87

Qy     57 --RRRGRETDSARRRRSRSRSLGNSVYSSPYDAGTPGTPGTPVATPVYATPVGTPMGTPS 114
Db     88 DSQRSARVQDVGRIMRRAQRSDVTDSVSSP--------------------------- 117

Qy    115 FHRGTPQYKQRSELGSQGKPLHRRRRSQSREPGHRSPSREPSADGRPSESAEPDDTLGGE 174
Db    118 ---------QRS----------RRYFTQGRGNGPSNLNSSTSAQ-FSTDPAEPNDE---P 154

Qy    175 YAYVWGTNVNIPDVLRAIRRFL----HNYRSSA---------HDLNSKYIQIIEETVERE 221
Db    155 VRVIWGTNVSIQECSNIFRDFLLSFKYKYRRELEEQAIEPEDHEL--YYVNQLNNIIELG 212

Qy    222 EDTLNIDMSDI--YDHDPDLYAKIVRYPLDIIPLLDTECQEV-------------ATSL 265
Db    213 LTNLNLDAKNLLSYPSTRKLYYQLINYPQEIIPIMDHTIKDCLIQIINDANATTSPAQSK 272

Qy    266 LPTFEKHI-EARPFNLK-ASVHMRELNPSDIDKLVSVKGMVIRCSSIIPEIKGAFFKCLV 323
Db    273 LDEIETNVYTIRPYNVNLVEKGIRELNPNDIDKLVSVKGLTLRSTSIIPDMKVAFFRCNA 332

Qy    324 CGHSPPLVTVVKGRVEEPTRCEKPECAARNAMSLIHNRCTFANKQIVRLQETPDAIPEGE 383
Db    333 CGHTVG-VEIDRGVISEPTKCPREVCGQTNSMVLIHNRSSFSDKQVIKLQETPDLVPDGQ 391

Qy    384 TPHTVSMCLYNTMVDAVKPGDRIEVTGVFKAMAVRVGPNQRTLRALYKTYIDCVHVKKSD 443
Db    392 TPHSINLCVYDELVDSCRAGDRVEVCGIFRSTPVRANPRQRALKNLYKTYLDIVHVKKID 451

Qy    444 RGRLQTE------DPMEMDKENDMYAGYHESDTSEAANEAKIQKLKELSKLPGIYDRLSR 497
Db    452 KRRLGGDVTTLEHELAEKDQE------VEQVRKITAEEEAKI---KEISERDDLYEILAR 502

Qy    498 SLAPSIWELEDIKKGLLCQLFGGKAKKIPSGASFRGDINVLLVGDPGTSKSQLLQYVHKI 557
Db    503 SLAPSIYEMDDVKKGILLQLFGGTNKTFTKGGRYRGDINILLCGDPSTSKSQILQYVHKI 562

Qy    558 APRGIYTSGRGSSAVGLTAYVTKDPETRETVLESGALVLSDRGICCIDEFDKMSDNARSM 617
Db    563 APRGVYTSGKGSSAVGLTAYITRDIDTKQLVLESGALVLSDGGVCCIDEFDKMSDATRSV 622

Qy    618 LHEVMEQQTVSVAKGGIIASLNARTSVLACANPSGSRYNARLSVIDNIQLPPTLLSRFDL 677
Db    623 LHEVMEQQTISIAKAGIITTLNARTSILASANPINSRYDPNLPVTGNIDLPPPLLSRFDL 682

Qy    678 IYLMLDKPDEQNDRRLARHLVALHYEN--YEVSKQDALDLQTLTAYITYARQHVHPTLSD 735
Db    683 VYLILDKVDESIDRQLARHLTDMYLEDAPETVNANSVLPVELLTLYIQYAKENFNPVMTE 742

Qy    736 EAAEDLINGYVEMRQKGNFPGSSKKVITATPRQLESMIRISEALARMRFSEVVEKVDAAE 795
Db    743 EGKNELVRSYVEMRKLGEDARSSEKRITATTRQLESMIRLSEAHAKMRLSERVELIDVKE 802

Qy    796 AVRLLDVALQQSATDHATGTIDMDLITTGVSASERIRRANLLAALRELIADKISPGSSSG 855
Db    803 AVRLIKSAIKDYATDPVTGRIDMDMIQTGTTAQQRRVQEDLVSEIMKIIEE-----NNNL 857

Qy    856 LKTSQLLEDIRSQSSVDVSLQDIKNALGSLQGEGFLTVHGDIVKR 900
Db    858 IRFNDLSVKLNERSSFRVENSVINEGLRRLQQEGKIMETGDSHRR 902
```

```
RESULT 14
US-11-097-143-1689
; Sequence 1689, Application US/11097143
; Publication No. US20050208558A1
; GENERAL INFORMATION:
;  APPLICANT: Venter, J. Craig
;  APPLICANT:  et al.
;  TITLE OF INVENTION: DETECTION KIT, SUCH AS NUCLEIC ACID
;  TITLE OF INVENTION:  ARRAYS, FOR DETECTING EXPRESSION OF 10,000 OR MORE
;  TITLE OF INVENTION:  DROSOPHILA GENES.
;  FILE REFERENCE: CL000728
;  CURRENT APPLICATION NUMBER: US/11/097,143
;  CURRENT FILING DATE:  2005-04-04
;  PRIOR APPLICATION NUMBER: 60/157,832
;  PRIOR FILING DATE: 1999-10-05
;  PRIOR APPLICATION NUMBER: 60/160,191
;  PRIOR FILING DATE: 1999-10-19
;  PRIOR APPLICATION NUMBER: 60/161,932
;  PRIOR FILING DATE: 1999-10-28
;  PRIOR APPLICATION NUMBER: 60/164,769
;  PRIOR FILING DATE: 1999-11-12
;  PRIOR APPLICATION NUMBER: 60/173,383
;  PRIOR FILING DATE: 1999-12-28
;  PRIOR APPLICATION NUMBER: 60/175,693
;  PRIOR FILING DATE: 2000-01-12
;  PRIOR APPLICATION NUMBER: 60/184,831
;  PRIOR FILING DATE: 2000-02-24
;  PRIOR APPLICATION NUMBER: 60/191,637
;  PRIOR FILING DATE: 2000-03-23
;  NUMBER OF SEQ ID NOS: 43008
;  SOFTWARE: FastSEQ for Windows Version 4.0
; SEQ ID NO 1689
;   LENGTH: 866
;   TYPE: PRT
;   ORGANISM: DROSOPHILA
US-11-097-143-1689

  Query Match             36.6%;  Score 1686.5;  DB 6;  Length 866;
  Best Local Similarity   41.6%;  Pred. No. 2e-117;
  Matches  386;  Conservative  161;  Mismatches  280;  Indels  101;  Gaps   20;

Qy         12 VSSPYPSQSEGVSTPLPQVTSPSFDNAASPVAGRRAVRQTPTSAVRRRGRET-------D 64
                 :|||  | | |  :||     :|:    |:   |      :|  |  |||       :     :
Db          1 MSSPARSPSVGGATPKQGARTPTRGIASQDV--ETPMRMGPGRAVRPSDNISLPPTSPGN 58

Qy         65 SARRRRSRSRSLGNSV----YSSPYDAGTPGTPGTPVATPVYATPVGTPMGTPSFHRGTP 120
                 :      | :|  || ::          ||| :  ||| :  |: :  ||
Db         59 ISLPATSPARGLGANMSEIDLSSPLNYGTPSSMGS-IRTP-------------------- 97

Qy        121 QYKQRSELGSQGKPLHRRRRSQSREPGHRSPSR------------EPSADGRPSESAEP- 167
                     ||   | :|  ||   |         |   |:   |              ||   :  : ||:  :|
Db         98 ----RS--GIRGTPLRAR-------PDIRTDKRIRQVAIGGGSGLEPIPE-KGSETTDPV 143

Qy        168 DDTLGGEYAYVWGTNVNIPDVLRAIRRFLHNY---------RSSAHDLNSK-YIQIIEET 217
                :: :        ||||||  :           :  |:   :                |        |:|     |:|  :||
Db        144 SESSQAPQLVVWGTNVVVSQCKSKFKSFIMRFIDPSAEQDEISENIDVNQPLYLQKLEEI 203

Qy        218 VEREEDTLNIDMSDIYDHDPDLYAKIVRYPLDIIPLLDTECQEVATSLLPT--FEKHIEA 275
                  ||   ||: :  :       | |||   :::  ||  ::||     |          |:            |          |       |:
Db        204 HTLEEPYLNLNCAHLKTFDQDLYRQLICYPQEVIPGFDMAINEMFFERYPAALLEHQIQV 263

Qy        276 RPFNLKASVHMRELNPSDIDKLVSVKGMVIRCSSIIPEIKGAFFKCLVCGHSPPLVTVVK 335
                |||       :  :||   |||    |:|:|:|:   ||||    |::||||::    |||   |    :|      |           |       |   :
Db        264 RPFNADKTRNMRSLNPEDMDQLISISGMVIRSSNVIPEMREAFFSCNICSFS-TTVEVDR 322

Qy        336 GRVEEPTRCEKPECAARNAMSLIHNRCTFANKQIVRLQETPDAIPEGETPHTVSMCLYNT 395
                ||:  :||  |        |        :        |||||     |   :||:|:|||:||   :         |:|||   |    :    :|
Db        323 GRINQPTLC--TNCNTNHCFRLIHNRSEFTDKQLVKLQESPDDMAAGQTPHNVLLYAHND 380

Qy        396 MVDAVKPGDRIEVTGVFKAMAVRVGPNQRTLRALYKTYIDCVHVKKSDRGRLQTEDPMEM 455
                 :||  |:||||: |||:::|    :: |           ::::|||:: |   ||  :|   |     ||   ||
Db        381 LVDKVQPGDRVTVTGIYRATPLKTGGLSSSVKSVYKTHVDVVHFRKVDNKRLY-ED---- 435

Qy        456 DKENDMYAGYHESDTSEAANEAKIQKLKELSKLPGIYDRLSRSLAPSIWELEDIKKGLLC 515
                               |                     ::: |: |:| |  ||||:|::|||:|  :|||||:|
Db        436 -----------EEGKDHIFPPERVELLQLLAKKPDIYDRLARAIAPSIYENDDIKKGILL 484

Qy        516 QLFGGKAKKIPS--GASFRGDINVLLVGDPGTSKSQLLQYVHKIAPRGIYTSGRGSSAVG 573
                |||||  ||    :        :||  :|::||   |||||||||:||||    :   ||    |||||||||||
Db        485 QLFGGTKKKHATLGRQNFRSEIHLLLCGDPGTSKSQMLQYVFNLVPRSQYTSGRGSSAVG 544

Qy        574 LTAYVTKDPETRETVLESGALVLSDRGICCIDEFDKMSDNARSMLHEVMEQQTVSVAKGG 633
                ||||||||||||: ||::|||||:|  |:|||||||||||:|: ||:|||||||||||:|:||  |
Db        545 LTAYVTKDPETRQLVLQTGALVLADNGVCCIDEFDKMNDSTRSVLHEVMEQQTLSIAKAG 604

Qy        634 IIASLNARTSVLACANPSGSRYNARLSVIDNIQLPPTLLSRFDLIYLMLDKPDEQNDRRL 693
                ||    ||||||:||  |||:   |::|   |   ::|||:|||   ||||||||||:|:||    ||      |:||
Db        605 IICQLNARTSILAAANPAESQWNKRKNIIDNVQLPHTLLSRFDLIFLVLDPQDEIFDKRL 664

Qy        694 ARHLVALHYENYEVSKQDALDLQTLTAYITYARQHVHPTLSDEAAEDLINGYVEMRQKGN 753
                 | |||:|:|            :             |:     |        ||  ||  |||:|:  ||||||  :  ||      ||:|:|:
Db        665 ASHLVSLYYVTRHEEEDTMFDMSVLRDYIAYAREHLSPTLSDEAQQRLIQAYVDMRK--- 721

Qy        754 FPGSSKKVITATPRQLESMIRISEALARMRFSEVVEKVDAAEAVRLLDVALQQSATDHAT 813
                    |:  :      |:|   ||||||:||:|||  |::|   |          ||  :|     ||   ||          ||:||||||     :
Db        722 -VGAGRGQISAYPRQLESLIRLSEAHAKVRLSNQVELLDVEEAWRLHREALKQSATDPLS 780

Qy        814 GTIDMDLITTGVSASERIRRANLLAALRELIADKISPGSSSGLKTSQLLEDIRSQSSVDV 873
                | ||: ::|||:|  :  |  :||:|:||::|  :     |            |           :         :|       ||:     |  :  :
Db        781 GKIDVGILTTGLSTAARKKRADLVAAIKENLKKK---GKVLTVPYQKLFSDIKEGSQIMI 837

Qy        874 SLQDIKNALGSLQGEGFLTVHGDIVKRV 901
                : :   ::||   :|  ||  :  |   |        |:
Db        838 TREQFEDALKEVQDEGAIVVMGKNTIRI 865


RESULT 15
US-10-425-114-68963
; Sequence 68963, Application US/10425114
; Publication No. US20040034888A1
; GENERAL INFORMATION:
;  APPLICANT: Liu, Jingdong
;  APPLICANT:  Zhou, Yihua
;  APPLICANT:  Kovalic, David K.
;  APPLICANT:  Screen, Steven E
;  APPLICANT:  Tabaska, Jack E
;  APPLICANT:  Cao, Yongwei
;  TITLE OF INVENTION: Nucleic Acid Molecules and Other Molecules Associated With
;  TITLE OF INVENTION:  Plants and Uses Thereof for Plant Improvement
;  FILE REFERENCE: 38-21(53313)B
;  CURRENT APPLICATION NUMBER: US/10/425,114
;  CURRENT FILING DATE:  2003-04-28
;  NUMBER OF SEQ ID NOS: 73128
; SEQ ID NO 68963
;   LENGTH: 456
;   TYPE: PRT
;   ORGANISM: Zea mays
;   FEATURE:
;   OTHER INFORMATION: Clone ID: LIB3150-112-C5_FLI.pep
US-10-425-114-68963

  Query Match             36.2%;  Score 1670.5;  DB 4;  Length 456;
  Best Local Similarity   70.4%;  Pred. No. 1.2e-116;
  Matches  326;  Conservative   64;  Mismatches   66;  Indels    7;  Gaps    2;

Qy        439 VKKSDRGRLQTEDPMEMDKENDMYAGYHESDTSEAANEAKIQKLKELSKLPGIYDRLSRS 498
                :||:|: ||  || |::|  |                    | |         |::|||||||||  ||:||:||
Db          1 IKKTDKSRLHVEDTMDIDNSN------ASKSTEEDFLSDKVEKLKELSKLPDIYERLTRS 54

Qy        499 LAPSIWELEDIKKGLLCQLFGGKAKKIPSGASFRGDINVLLVGDPGTSKSQLLQYVHKIA 558
                |||:||||:|:|:|||||||||      |:|||||||||||:|||||||||||||||||:|||::
Db         55 LAPNIWELDDVKRGLLCQLFGGNPLKLPSGASFRGDINILLVGDPGTSKSQLLQYMHKLS 114

Qy        559 PRGIYTSGRGSSAVGLTAYVTKDPETRETVLESGALVLSDRGICCIDEFDKMSDNARSML 618
```

```
                 |||||||||||||||||||||||||| |||||||||||||:|:||:|||||||||||||
Db        115 PRGIYTSGRGSSAVGLTAYVTKDPETGETVLESGALVLSDKGVCCVDEFDKMSDNARSML 174

Qy        619 HEVMEQQTVSVAKGGIIASLNARTSVLACANPSGSRYNARLSVIDNIQLPPTLLSRFDLI 678
              ||||||||||:|| ||||||||||||||||||: |||| |||||||| | ||||||||||
Db        175 HEVMEQQTVSIAKAGIIASLNARTSVLACANPTESRYNPRLSVIDNIHLAPTLLSRFDLI 234

Qy        679 YLMLDKPDEQNDRRLARHLVALHYENYEVSKQDALDLQTLTAYITYARQHVHPTLSDEAA 738
              ||:||| ||| |||||:|:|:||:||  : : : |||||| :||:|||::: | ||||||
Db        235 YLILDKADEQTDRRLAKHIVSLHFENPNLEELEVLDLQTLVSYISYARKYIQPQLSDEAA 294

Qy        739 EDLINGYVEMRQKGNFPGSSKKVITATPRQLESMIRISEALARMRFSEVVEKVDAAEAVR 798
              |:|  ||||||::|| ||| ||||||| ||:||:||:||||||||||||||  |  || |
Db        295 EELTRGYVEMRKRGNSPGSRKKVITATARQIESLIRLSEALARMRFSEVVEVRDVVEAFR 354

Qy        799 LLDVALQQSATDHATGTIDMDLITTGVSASERIRRANLLAALRELIADKISPGSSSGLKT 858
              ||:||:||||||||||||||||| ||:||||| || ||:|| | |||:|:  |  |  :: 
Db        355 LLEVAMQQSATDHATGTIDMDLIMTGISASERQRRENLVAATRNLIAEKMQLGGPS-MRM 413

Qy        859 SQLLEDIRSQSSVDVSLQDIKNALGSLQGEGFLTVHGDIVKRV 901
               :|||::| |||::: : ::: |||:|  || : :||| |:||
Db        414 IELLEELRKQSSMEIHMHELRGALGTLMTEGAVVIHGDNVRRV 456
```

Search completed: December 7, 2005, 17:47:33
Job time : 175 secs

```
OM protein - protein search, using sw model

Run on:          December  7, 2005, 17:28:10 ; Search time 239 Seconds
                                              (without alignments)
                                              2659.753 Million cell updates/sec

Title:           US-10-768-511-8
Perfect score:   4609
Sequence:        1 MENNDALDIGAVSSPYPSQS..........LGSLQGEGFLTVHGDIVKRV 901

Scoring table:   BLOSUM62
                 Gapop 10.0 , Gapext 0.5

Searched:        2166443 seqs, 705528306 residues

Total number of hits satisfying chosen parameters:      2166443

Minimum DB seq length: 0
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%
                 Maximum Match 100%
                 Listing first 45 summaries

Database :        UniProt_05.80:*
                 1:  uniprot_sprot:*
                 2:  uniprot_trembl:*

      Pred. No. is the number of results predicted by chance to have a
      score greater than or equal to the score of the result being printed,
      and is derived by analysis of the total score distribution.
```

SUMMARIES

Result No.	Score	% Query Match	Length	DB	ID	Description
1	2537	55.0	911	2	Q5JKB0_ORYSA	Q5jkb0 oryza sativ
2	2332	50.6	720	2	Q9SIV8_ARATH	Q9siv8 arabidopsis
3	1909	41.4	886	2	Q86IF1_DICDI	Q86if1 dictyosteli
4	1890.5	41.0	1020	2	Q4PE17_USTMA	Q4pe17 ustilago ma
5	1884	40.9	1023	2	Q4X1R4_ASPFU	Q4x1r4 aspergillus
6	1873	40.6	1013	2	Q7SHS5_NEUCR	Q7shs5 neurospora
7	1870	40.6	931	1	MCM4_SCHPO	P29458 schizosacch
8	1856.5	40.3	989	2	Q55I48_CRYNE	Q55148 cryptococcu
9	1856.5	40.3	989	2	Q5K7N5_CRYNE	Q5k7n5 cryptococcu
10	1833	39.8	933	1	CDC54_YEAST	P30665 saccharomyc
11	1833	39.8	1020	2	Q4IQK4_GIBZE	Q4iqk4 gibberella
12	1821	39.5	862	1	MCM4_MOUSE	P49717 mus musculu
13	1821	39.5	862	2	Q542F4_MOUSE	Q542f4 mus musculu
14	1817.5	39.4	863	2	Q6GL41_XENTR	Q6gl41 xenopus tro
15	1817	39.4	862	2	Q8C1Z0_MOUSE	Q8c1z0 mus musculu
16	1814.5	39.4	924	2	Q6FN63_CANGA	Q6fn63 candida gla
17	1811	39.3	862	2	Q921D5_MOUSE	Q921d5 mus musculu
18	1810.5	39.3	863	1	MCM4_HUMAN	P33991 homo sapien
19	1809.5	39.3	863	2	Q6GQ77_XENLA	Q6gq77 xenopus lae
20	1808.5	39.2	858	2	Q5XK83_XENLA	Q5xk83 xenopus lae
21	1807	39.2	862	2	Q9D077_MOUSE	Q9d077 mus musculu
22	1804.5	39.2	911	2	Q6BP39_DEBHA	Q6bp39 debaryomyce
23	1801.5	39.1	863	1	MCM4_XENLA	P30664 xenopus lae
24	1793.5	38.9	892	2	Q6CSV7_KLULA	Q6csv7 kluyveromyc
25	1791.5	38.9	845	2	Q6NZV2_BRARE	Q6nzv2 brachydanio
26	1785	38.7	888	2	Q75AE3_ASHGO	Q75ae3 ashbya goss
27	1784.5	38.7	858	2	O42589_XENLA	O42589 xenopus lae
28	1781	38.6	924	2	Q6C2V4_YARLI	Q6c2v4 yarrowia li
29	1754	38.1	910	2	Q59M26_CANAL	Q59m26 candida alb
30	1754	38.1	912	2	Q59M39_CANAL	Q59m39 candida alb
31	1690.5	36.7	677	2	Q5DTS8_MOUSE	Q5dts8 mus musculu
32	1686.5	36.6	866	1	MCM4_DROME	Q26454 drosophila
33	1684.5	36.5	876	2	Q7PPI5_ANOGA	Q7ppi5 anopheles g
34	1613	35.0	750	2	Q6PHK9_BRARE	Q6phk9 brachydanio
35	1565	34.0	817	2	Q61XQ3_CAEBR	Q61xq3 caenorhabdi
36	1562.5	33.9	823	2	Q95XQ8_CAEEL	Q95xq8 caenorhabdi
37	1542.5	33.5	608	2	Q50Q32_ENTHI	Q50q32 entamoeba h
38	1514.5	32.9	896	2	Q5CTW9_CRYPV	Q5ctw9 cryptospori
39	1503.5	32.6	894	2	Q5CGN7_CRYHO	Q5cgn7 cryptospori
40	1490	32.3	632	2	Q525B3_MAGGR	Q525b3 magnaporthe
41	1446.5	31.4	708	2	Q8SSE5_ENCCU	Q8sse5 encephalito
42	1356.5	29.4	556	2	Q5B060_EMENI	Q5b060 aspergillus
43	1318.5	28.6	1005	2	Q8IEE5_PLAF7	Q8iee5 plasmodium
44	1310.5	28.4	944	2	Q7RJ52_PLAYO	Q7rj52 plasmodium
45	1309.5	28.4	1005	2	Q9GR06_PLAFA	Q9gr06 plasmodium

ALIGNMENTS

```
RESULT 1
Q5JKB0_ORYSA
ID   Q5JKB0_ORYSA PRELIMINARY;      PRT;   911 AA.
AC   Q5JKB0;
DT   10-MAY-2005 (TrEMBLrel. 30, Created)
DT   10-MAY-2005 (TrEMBLrel. 30, Last sequence update)
DT   10-MAY-2005 (TrEMBLrel. 30, Last annotation update)
DE   Putative replication licensing factor MCM4.
GN   Name=OSJNBa0051H17.26;
OS   Oryza sativa (japonica cultivar-group).
OC   Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
OC   Spermatophyta; Magnoliophyta; Liliopsida; Poales; Poaceae;
OC   Ehrhartoideae; Oryzeae; Oryza.
OX   NCBI_TaxID=39947;
RN   [1]
RP   NUCLEOTIDE SEQUENCE.
RA   Sasaki T., Matsumoto T., Yamamoto K., Sakata K., Baba T., Katayose Y.,
RA   Wu J., Niimura Y., Cheng Z., Nagamura Y., Antonio B.A., Kanamori H.,
RA   Hosokawa S., Masukawa M., Arikawa K., Chiden Y., Hayashi M.,
RA   Okamoto M., Ando T., Aoki H., Arita K., Hamada M., Harada C.,
RA   Hijishita S., Honda M., Ichikawa Y., Idonuma A., Iijima M., Ikeda M.,
RA   Ikeno M., Itoh S., Itoh T., Itoh Y., Itoh Y., Iwabuchi A., Kamiya K.,
RA   Karasawa W., Katagiri S., Kikuta A., Kobayashi N., Kono I.,
RA   Machita K., Maehara T., Mizuno H., Mizubayashi T., Mukai Y.,
RA   Nagasaki H., Nakashima M., Nakama Y., Nakamichi Y., Nakamura M.,
RA   Namiki N., Negishi M., Ohta I., Ono N., Saji S., Sakai K., Shibata M.,
RA   Shimokawa T., Shomura A., Song J., Takazaki Y., Terasawa K., Tsuji K.,
RA   Waki K., Yamagata H., Yamane H., Yoshiki S., Yoshihara R., Yukawa K.,
RA   Zhong H., Iwama H., Endo T., Ito H., Hahn J.H., Kim H.I., Eun M.Y.,
RA   Yano M., Jiang J., Gojobori T.;
RT   "The genome sequence and structure of rice chromosome 1.";
RL   Nature 420:312-316(2002).
DR   EMBL; AP004232; BAD88098.1; -; Genomic_DNA.
DR   GO; GO:0005634; C:nucleus; IEA.
DR   GO; GO:0005524; F:ATP binding; IEA.
DR   GO; GO:0003677; F:DNA binding; IEA.
DR   GO; GO:0008094; F:DNA-dependent ATPase activity; IEA.
DR   GO; GO:0017111; F:nucleoside-triphosphatase activity; IEA.
DR   GO; GO:0000166; F:nucleotide binding; IEA.
DR   GO; GO:0006270; P:DNA replication initiation; IEA.
DR   InterPro; IPR003593; AAA_ATPase.
DR   InterPro; IPR001208; MCM.
DR   InterPro; IPR008047; MCM_4.
DR   Pfam; PF00493; MCM; 1.
DR   PRINTS; PR01657; MCMFAMILY.
DR   PRINTS; PR01660; MCMPROTEIN4.
DR   ProDom; PD001041; MCM; 1.
DR   SMART; SM00382; AAA; 1.
DR   SMART; SM00350; MCM; 1.
DR   PROSITE; PS00847; MCM_1; 1.
DR   PROSITE; PS50051; MCM_2; 1.
SQ   SEQUENCE   911 AA;  100417 MW;  94B01E5DD73C1405 CRC64;

  Query Match           55.0%;  Score 2537;  DB 2;  Length 911;
```

```
  Best Local Similarity   57.2%;  Pred. No. 4.8e-132;
  Matches  514;  Conservative  137;  Mismatches  176;  Indels   72;  Gaps   16;

Qy         24 STPLPQVTSPSFDNAASPVAGRRAVRQTPTSAVRRRGRETDSARRRRSRSRSLGNSVYSS 83
              ::| | |:||     :||:     |    :|  : || |       ||||          :|
Db         13 NSPPPSVSSPDV-RPSSPLP---ATNSSPPQSGRRGG-----GRRRRGS---------AS 54

Qy         84 PYDAGTPGTPGTPVATPVYATPVGTPMGTPSFHRGTPQYKQRSELGSQGKPLHRRRRSQS 143
              || : :|     |         : ||        |||       | :| :   | :  |
Db         55 PYPS-SPSLGG-------FETPPHPGRRTPSGGAAARQQRQ-NWTGGRFPP--------- 96

Qy        144 REPGHRSPSREPSADGRP--SESAEPD--DTLGG---------EYAYVWGTNVNIPDVLR 190
                    :||    ||  |    || : |  :| ||             :|||||::: ||
Db         97 ------TPSTPMSTDDVPLSSEAGDEDTPETDGGGGGGAGADATPVFVWGTNISVQDVNA 150

Qy        191 AIRRFLHNYRSSAHDL--------NSKYIQIIEETVERE-EDTLNIDMSDIYDHDPDLYA 241
              || ||| ::|      |                ||:: |    :| |   ::|:::  |::||||||
Db        151 AILRFLRHFR-DPRDAGRVDPVMDEGKYMRAIHRILELEGGESLDVNAHDVFDHDPDLYG 209

Qy        242 KIVRYPLDIIPLLDTECQEVATSLLPTFEKHIEARPFNLKASVHMRELNPSDIDKLVSVK 301
              |:|||||::: : |      ::    : | |||||: | :|||:|| :| ||||||:|:||:|
Db        210 KMVRYPLEVLAIFDIVLMDLVARIEPLFEKHIQTRIYNLKSSVCLRNLNPSDIEKMVSIK 269

Qy        302 GMVIRCSSIIPEIKGAFFKCLVCGHSPPLVTVVKGRVEEPTRCEKPECAARNAMSLIHNR 361
              ||:|||||:|||:| | |:|||||        | | :||| ||  |:| :| | |:|:|:|||
Db        270 GMIIRCSSVIPELKEAVFRCLVCGFYSEPVMVDRGRVTEPHICQKEQCKATNSMTLVHNR 329

Qy        362 CTFANKQIVRLQETPDAIPEGETPHTVSMCLYNTMVDAVKPGDRIEVTGVFKAMAVRVGP 421
              | ||:||:|:||||| ||| ||||||  |||||:|:|||:|::|:||:|::|:|||
Db        330 CRFADKQIIKLQETPDEIPEGGTPHTVSVLMHDKLVDAGKPGDRVEITGIYRAMSIRVGP 389

Qy        422 NQRTLRALYKTYIDCVHVKKSDRGRLQTEDPMEMDKENDMYAGYHESDTSEAANEAKIQK 481
               ||:::::||||||:|:||:|:  ||   || || |  |                   | :    |::|
Db        390 TQRTVKSIFKTYIDCLHIKKTDKSRLHVEDSMETDNPN------ANKTTEDDFLRDKVEK 443

Qy        482 LKELSKLPGIYDRLSRSLAPSIWELEDIKKGLLCQLFGGKAKKIPSGASFRGDINVLLVG 541
              ||||||| ||||:||||:||||:|:|:|:||||||||  |  ::||||||||||||:||||
Db        444 LKELSKLPDIYDRLTRSLAPNIWELDDVKRGLLCQLFGGNALRLPSGASFRGDINILLVG 503

Qy        542 DPGTSKSQLLQYVHKIAPRGIYTSGRGSSAVGLTAYVTKDPETRETVLESGALVLSDRGI 601
              |||||||||||:||::|||||||||||||||||||||||||||||  ||||||||||||:|:
Db        504 DPGTSKSQLLQYMHKLSPRGIYTSGRGSSAVGLTAYVTKDPETGETVLESGALVLSDKGV 563

Qy        602 CCIDEFDKMSDNARSMLHEVMEQQTVSVAKGGIIASLNARTSVLACANPSGSRYNARLSV 661
              ||||||||||||||||||||||||||||:||:|||||||||||||||||| |||| |||
Db        564 CCIDEFDKMSDNARSMLHEVMEQQTVSIAKAGIIASLNARTSVLACANPTESRYNPRLSV 623

Qy        662 IDNIQLPPTLLSRFDLIYLMLDKPDEQNDRRLARHLVALHYENYEVSKQDALDLQTLTAY 721
              |||| |||||||||||||:|||  |||  ||||:|:|:||:|    : : :  ||| || ||
Db        624 IDNIHLPPTLLSRFDLIYLILDKADEQTDRRLAKHIVSLHFENPNIEELEVLDLPTLVAY 683

Qy        722 ITYARQHVHPTLSDEAAEDLINGYVEMRQKGNFPGSSKKVITATPRQLESMIRISEALAR 781
              |:|||:|: | ||||||:|   |||||::|| ||| |||||| ||:||:|:|:||||||
Db        684 ISYARKHIQPQLSDEAAEELTRGYVEMRKRGNSPGSRKKVITATARQIESLIRLSEALAR 743

Qy        782 MRFSEVVEKVDAAEAVRLLDVALQQSATDHATGTIDMDLITTGVSASERIRRANLLAALR 841
              |||||:||   |  ||  |||:||:|:||||||||||||||| ||:||||| || ||:|| |
Db        744 MRFSEMVEVQDVVEAFRLLEVAMQQSATDHATGTIDMDLIMTGISASERQRRDNLVAATR 803

Qy        842 ELIADKISPGSSSGLKTSQLLEDIRSQSSVDVSLQDIKNALGSLQGEGFLTVHGDIVKR 900
                |: :|:  |   | ::  :||:|:|| |||::| | |:: |||:|   || : :||| |||
Db        804 NLVMEKMQLGGPS-VRMIELLEEIRKQSSMEVHLHDLRGALGTLMTEGAVVIHGDSVKR 861

RESULT 2
Q9SIV8_ARATH
ID   Q9SIV8_ARATH PRELIMINARY;      PRT;   720 AA.
AC   Q9SIV8;
DT   01-MAY-2000 (TrEMBLrel. 13, Created)
DT   01-MAY-2000 (TrEMBLrel. 13, Last sequence update)
DT   01-OCT-2003 (TrEMBLrel. 25, Last annotation update)
DE   Putative CDC21 protein.
GN   Name=At2g16440;
OS   Arabidopsis thaliana (Mouse-ear cress).
OC   Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
OC   Spermatophyta; Magnoliophyta; eudicotyledons; core eudicotyledons;
OC   rosids; eurosids II; Brassicales; Brassicaceae; Arabidopsis.
OX   NCBI_TaxID=3702;
RN   [1]
RP   NUCLEOTIDE SEQUENCE.
RA   Lin X., Kaul S., Shea T.P., Fujii C.Y., Shen M., VanAken S.E.,
RA   Barnstead M.E., Mason T.M., Bowman C.L., Ronning C.M., Benito M.-I.,
RA   Carrera A.J., Creasy T.H., Buell C.R., Town C.D., Nierman W.C.,
RA   Fraser C.M., Venter J.C.;
RL   Submitted (MAR-2000) to the EMBL/GenBank/DDBJ databases.
RN   [2]
RP   NUCLEOTIDE SEQUENCE.
RA   Town C.D., Kaul S.;
RL   Submitted (FEB-2002) to the EMBL/GenBank/DDBJ databases.
DR   EMBL; AC007047; AAD22296.1; -; Genomic_DNA.
DR   PIR; C84540; C84540.
DR   GO; GO:0005634; C:nucleus; IEA.
DR   GO; GO:0005524; F:ATP binding; IEA.
DR   GO; GO:0003677; F:DNA binding; IEA.
DR   GO; GO:0008094; F:DNA-dependent ATPase activity; IEA.
DR   GO; GO:0017111; F:nucleoside-triphosphatase activity; IEA.
DR   GO; GO:0000166; F:nucleotide binding; IEA.
DR   GO; GO:0006270; P:DNA replication initiation; IEA.
DR   InterPro; IPR003593; AAA_ATPase.
DR   InterPro; IPR001208; MCM.
DR   InterPro; IPR008047; MCM_4.
DR   Pfam; PF00493; MCM; 1.
DR   PRINTS; PR01657; MCMFAMILY.
DR   PRINTS; PR01660; MCMPROTEIN4.
DR   ProDom; PD001041; MCM; 1.
DR   SMART; SM00382; AAA; 1.
DR   SMART; SM00350; MCM; 1.
DR   PROSITE; PS00847; MCM_1; 1.
DR   PROSITE; PS50051; MCM_2; 1.
SQ   SEQUENCE   720 AA;  80950 MW;  1128F80D8AA5B018 CRC64;

  Query Match          50.6%;  Score 2332;  DB 2;  Length 720;
  Best Local Similarity   61.0%;  Pred. No. 8.1e-121;
  Matches  448;  Conservative  124;  Mismatches  135;  Indels   28;  Gaps    6;

Qy        154 EPSADGRPSESAEPDDTLGGEYAYVWGTNVNIPDVLRAIRRFLHNY---RSSAHDL--NS 208
              ||       |      |||         :|||||::: ||  ||  |: ::    | :: ||
Db          5 EPLPSSDDGEEDGGDDTTP---TFVWGTNISVQDVKSAIEMFVKHFREARENSDDLFREG 61

Qy        209 KYIQIIEETVEREEDTLNIDMSDIYDHDPDLYAKIVRYPLDIIPLLDTECQEVATSLLPT 268
              ||:  | : :| | : :::|  |::|:|||| |:|||||::: : |      :: :::
Db         62 KYMVSIRKVIEIEGEWIDVDAFDVFDYDPDLYNKMVRYPLEVLAIFDIVLMDIVSTINRL 121

Qy        269 FEKHIEARPFNLKASVHMRELNPSDIDKLVSVKGMVIRCSSIIPEIKGAFFKCLVCGHSP 328
              ||||:: | |||: |    || ||||||:|::|:||:|| |||||||:  | |:||||:
Db        122 FEKHVQVRIFNLRTSTSMRNLNPSDIEKMISLKGMIIRSSSIIPEIREAVFRCLVCGYFS 181

Qy        329 PLVTVVKGRVEEPTRCEKPECAARNAMSLIHNRCTFANKQIVRLQETPDAIPEGETPHTV 388
                : | :|:: ||   | | ||   :|:|:|:||||  ||:||||||||||| |||| |||||
Db        182 DPIIVDRGKISEPPTCLKQECMTKNSMTLVHNRCRFADKQIVRLQETPDEIPEGGTPHTV 241

Qy        389 SMCLYNTMVDAVKPGDRIEVTGVFKAMAVRVGPNQRTLRALYKTYIDCVHVKKSDRGRLQ 448
              |: |:: :||  ||||||||||:::||  |||||    ||:::::|||||||:|:||: : |:
Db        242 SLLLHDKLVDNGKPGDRIEVTGIYRAMTVRVGPAHRTVKSVFKTYIDCLHIKKASKLRMS 301

Qy        449 TEDPMEMDKENDMYAGYHESDTSEAANEAKIQKLKELSKLPGIYDRLSRSLAPSIWELED 508
               ||||::              | |    :   ::| :||||  | ||:||||||||||:||||:|
Db        302 AEDPMDV-------------DNSLRRVDEDLRKFQELSKQPDIYERLSRSLAPNIWELDD 348

Qy        509 IKKGLLCQLFGGKAKKIPSGASFRGDINVLLVGDPGTSKSQLLQYVHKIAPRGIYTSGRG 568
              :||||||||||||  |    :  |||:|||||:||||||||||||||||||:|::|||||||||
Db        349 VKKGLLCQLFGGNALNLASGANFRGDINILLVGDPGTSKSQLLQYIHKLSPRGIYTSGRG 408

Qy        569 SSAVGLTAYVTKDPETRETVLESGALVLSDRGICCIDEFDKMSDNARSMLHEVMEQQTVS 628
              |||||||||| |||||  |||||||||||||||||||||||||||:|||||||||||||||
Db        409 SSAVGLTAYVAKDPETGETVLESGALVLSDRGICCIDEFDKMSDSARSMLHEVMEQQTVS 468
```

```
Qy        629 VAKGGIIASLNARTSVLACANPSGSRYNARLSVIDNIQLPPTLLSRFDLIYLMLDKPDEQ 688
              :|| ||||||||||||||||||||||| |||||:|| ||||||||||||||:|||||||
Db        469 IAKAGIIASLNARTSVLACANPSGSRYNPRLSVIENIHLPPTLLSRFDLIYLILDKPDEQ 528

Qy        689 NDRRLARHLVALHYENYEVSKQDALDLQTLTAYITYARQHVHPTLSDEAAEDLINGYVEM 748
               |||||:|:||||:|| | ::::|:|: ||| |::|||:::|| |||||||:| ||||:
Db        529 TDRRLAKHIVALHFENAESAQEEAIDITTLTTYVSYARKNIHPKLSDEAAEELTRGYVEL 588

Qy        749 RQKGNFPGSSKKVITATPRQLESMIRISEALARMRFSEVVEKVDAAEAVRLLDVALQQSA 808
              |: | | ||||||||||||:||:||:||||||||||| ||| |   || ||| ||:||||
Db        589 RKAGKFAGSSKKVITATPRQIESLIRLSEALARMRFSEWVEKHDVDEAFRLLRVAMQQSA 648

Qy        809 TDHATGTIDMDLITTGVSASERIRRANLLAALRELIADKISPGSSSGLKTSQLLEDIRSQ 868
              |||||||||||| ||||||||:||     :::|:: :||     :   :   ||||:::
Db        649 TDHATGTIDMDLINTGVSASERMRRDTFASSIRDIALEKI----TMEMLLLQLLEELKKH 704

Qy        869 S---SVDVSLQDIKN 880
                  : :: | |: :
Db        705 GGNINTEIHLHDVSH 719


RESULT 3
Q86IF1_DICDI
ID   Q86IF1_DICDI PRELIMINARY;      PRT;   886 AA.
AC   Q86IF1;
DT   01-JUN-2003 (TrEMBLrel. 24, Created)
DT   01-JUN-2003 (TrEMBLrel. 24, Last sequence update)
DT   10-MAY-2005 (TrEMBLrel. 30, Last annotation update)
DE   Similar to essential for initiation of DNA replication; homolog of S.
DE   pombe CDC21; Cdc54p [Saccharomyces cerevisiae] (Hypothetical
DE   protein).
GN   ORFNames=DDB0167140;
OS   Dictyostelium discoideum (Slime mold).
OC   Eukaryota; Mycetozoa; Dictyosteliida; Dictyostelium.
OX   NCBI_TaxID=44689;
RN   [1]
RP   NUCLEOTIDE SEQUENCE.
RC   STRAIN=AX4;
RX   MEDLINE=22092622; PubMed=12097910; DOI=10.1038/nature00847;
RA   Gloeckner G., Eichinger L., Szafranski K., Pachebat J.A.,
RA   Bankier A.T., Dear P.H., Lehmann R., Baumgart C., Parra G.,
RA   Abril J.F., Guigo R., Kumpf K., Tunggal B., Cox E., Quail M.A.,
RA   Platzer M., Rosenthal A., Noegel A.A.;
RT   "Sequence and analysis of chromosome 2 of Dictyostelium discoideum.";
RL   Nature 418:79-85(2002).
RN   [2]
RP   NUCLEOTIDE SEQUENCE.
RC   STRAIN=AX4;
RA   Baumgart C.;
RL   Submitted (FEB-2004) to the EMBL/GenBank/DDBJ databases.
RN   [3]
RP   NUCLEOTIDE SEQUENCE.
RC   STRAIN=AX4;
RA   Eichinger L., Pachebat J.A., Gloeckner G., Rajandream M.-A.,
RA   Sucgang R., Berriman M., Song J., Olsen R., Szafranski K., Xu Q.,
RA   Tunggal B., Kummerfeld S., Madera M., Konfortov B.A., Rivero F.,
RA   Bankier A.T., Lehmann R., Hamlin N., Davies R., Gaudet P., Fey P.,
RA   Pilcher K., Chen G., Saunders D., Sodergren E., Davis P.,
RA   Kerhornou A., Nie X., Hall N., Anjard C., Hemphill L., Bason N.,
RA   Farbrother P., Desany B., Just E., Morio T., Rost R., Churcher C.,
RA   Cooper J., Haydock S., van Driessche N., Cronin A., Goodhead I.,
RA   Muzny D., Mourier T., Pain A., Lu M., Harper D., Lindsay R.,
RA   Hauser H., James K., Quiles M., Mohan M.B., Saito T., Buchrieser C.,
RA   Wardroper A., Felder M., Thangavelu M., Johnson D., Knights A.,
RA   Loulseged H., Mungall K., Oliver K., Price C., Quail M.A.,
RA   Urushihara H., Hernandez J., Rabbinowitsch E., Steffen D., Sanders M.,
RA   Ma J., Kohara Y., Sharp S., Simmonds M., Spiegler S., Tivey A.,
RA   Sugano S., White B., Walker D., Woodward J., Winckler T., Tanaka Y.,
RA   Shaulsky G., Schleicher M., Weinstock G., Rosenthal A., Cox E.C.,
RA   Chisholm R.L., Gibbs R., Loomis W.F., Platzer M., Kay R.R.,
RA   Williams J., Dear P.H., Noegel A.A., Barrell B., Kuspa A.;
RT   "The genome of the social amoeba Dictyostelium discoideum.";
RL   Nature 0:0-0(2005).
DR   EMBL; AC116551; AAO52157.1; -; Genomic_DNA.
DR   EMBL; AAFI01000030; EAL69562.1; -; Genomic_DNA.
DR   GO; GO:0005634; C:nucleus; IEA.
DR   GO; GO:0005524; F:ATP binding; IEA.
DR   GO; GO:0003677; F:DNA binding; IEA.
DR   GO; GO:0008094; F:DNA-dependent ATPase activity; IEA.
DR   GO; GO:0006270; P:DNA replication initiation; IEA.
DR   InterPro; IPR001208; MCM.
DR   InterPro; IPR008047; MCM_4.
DR   Pfam; PF00493; MCM; 1.
DR   PRINTS; PR01657; MCMFAMILY.
DR   PRINTS; PR01660; MCMPROTEIN4.
DR   ProDom; PD001041; MCM; 1.
DR   SMART; SM00350; MCM; 1.
DR   PROSITE; PS00847; MCM_1; 1.
DR   PROSITE; PS50051; MCM_2; 1.
KW   Hypothetical protein.
SQ   SEQUENCE   886 AA;  100648 MW;  DD545AA3ED9421DF CRC64;

  Query Match             41.4%;  Score 1909;  DB 2;  Length 886;
  Best Local Similarity   49.8%;  Pred. No. 3.2e-97;
  Matches  397;  Conservative  128;  Mismatches  174;  Indels   98;  Gaps   14;

Qy        179 WGTNVNIPDVLRAIRRFLHNY-------RSSAHDLNSK--------------------- 209
              ||| | : ||    |:|: |:        | : :|:|:
Db        103 WGTTVKVEDVKERFRKFIFNFPTPKERKRFNINDINNNNNNNNNNSEEMNDDDDYYYDE 162

Qy        210 ---------------YIQIIEETVEREEDTLNIDMSDIYDHDPDLYAKIVRYPLDIIPLL 254
                             | :::   | ::  |||:|  :    | :|| :  | :| ::||||
Db        163 EQDKLIKKPEKVQYLYRDLLKVLNETKKCHLNINMKYLVQFDMELYLQWVSFPNEMIPLL 222

Qy        255 DTECQEVATSLLPTFEKH------------------IEARPFNLKASVHMRELNPSDIDK 296
              || |  :   | |   ::                  ||  ||||      ||:||||||||
Db        223 DEEINLIYRELFPITPRNGDDEDDDDDDDEDEEDYRIELHPFNLLRKTPMRDLNPSDIDK 282

Qy        297 LVSVKGMVIRCSSIIPEIKGAFFKCLVCG---HSPPLVTVVKGRVEEPTRCEKPECAARN 353
              ::|: |::|| ||||||| ||| | ||    |:       | ||:::||: |     | ::
Db        283 IISISGLIIRSSSIIPEIKQAFFMCAVCEATFHA----NVEKGKIQEPSECS--NCKSKQ 336

Qy        354 AMSLIHNRCTFANKQIVRLQETPDAIPEGETPHTVSMCLYNTMVDAVKPGDRIEVTGVFK 413
              ::|:||||| | :|| ::||||||||||||||||:: |    ::|  |||||:|:||||
Db        337 SLSIIHNRCLFGDKQYIKLQETPDAIPEGETPHTVALFAYGDLIDIAKPGDRVELTGVFK 396

Qy        414 AMAVRVGPNQRTLRALYKTYIDCVHVKKSDRGRLQTEDPMEMD----------KENDMYA 463
              |  :| | | |:||::|||||| :|:|::|:|: :|  |              || :
Db        397 ASPMRAGSN-RSLRSIYKTYIDILHIKRTDKGKYDDDDDHDDNTGGGGTGTGKETNENL 455

Qy        464 GYHESDTSEAANEAKIQKLKELSKLPGIYDRLSRSLAPSIWELEDIKKGLLCQLFGGKAK 523
               : : |  : : | | ::: |||  ||  |||  :::|:|:||||||||||:||||||| |
Db        456 DFEDLDEFDLSEE-KEREIIELSKKPDIYDIVTKSIAPNIWELEDIKKGILCQLFGGSKK 514

Qy        524 KIPS-GASFRGDINVLLVGDPGTSKSQLLQYVHKIAPRGIYTSGRGSSAVGLTAYVTKDP 582
                   | ||||||:|| ||||||||||||| |||||||||||:||||||||||||:|||
Db        515 SYQDYGGKFRGDINILLCGDPGTSKSQLLSYVHKIAPRGIYTSGKGSSAVGLTAYITKDP 574

Qy        583 ETRETVLESGALVLSDRGICCIDEFDKMSDNARSMLHEVMEQQTVSVAKGGIIASLNART 642
              :|:||||||||||||:|:||||||||||:|  ||:||||||||||:|| ||| :||||
Db        575 DTKETVLESGALVLSDKGVCCIDEFDKMNDQTRSILHEVMEQQTVSIAKAGIICTLNART 634

Qy        643 SVLACANPSGSRYNARLSVIDNIQLPPTLLSRFDLIYLMLDKPDEQNDRRLARHLVALHY 702
              |:|| |||||||  :|||::||||||||||||||||:||| :|::||:|||||:::
Db        635 SILASANPSGSRYMPKLSVVENIQLPPTLLSRFDLIYLVLDKANERSDRQLARHLVSMYW 694

Qy        703 ENYEVSKQDALDLQTLTAYITYARQHVHPTLSDEAAEDLINGYVEMRQKGNFPGSSKKVI 762
               :  ||    :  :||| || |||:|::| |:|::|: |: ||:|||      ||||| |
Db        695 DETPVS-HFTIPKETLTNYIQYARKHINPKLTDDSAKCLVQGYLEMRSM----GSSKKTI 749

Qy        763 TATPRQLESMIRISEALARMRFSEVVEKVDAAEAVRLLDVALQQSATDHATGTIDMDLIT 822
              :||||||||:|||:||  :|   ||:|| ||||:|  |  |   |||||||||
Db        750 SATPRQLESLIRIAEAHARIRFSEFVEPLDVEEAIRLIKVALQQAAIDPENGTIDMDLIT 809

Qy        823 TGVSASERIRRANLLAALRELIADKISPGSSSGLKTSQLLEDIRSQSSVDV--SLQDIKN 880
```

```
Db        810 TGRSASSREAITRLKSHIKQKIGKK-------HLTLDQLLKLLTIQNQVQTIQQIEEIKE 862

Qy        881 ALGSLQGEGFLTVHGDI 897

Db        863 ALRQLQDEEIIQSSGGI 879


RESULT 4
Q4PE17_USTMA
ID   Q4PE17_USTMA PRELIMINARY;      PRT;  1020 AA.
AC   Q4PE17;
DT   13-SEP-2005 (TrEMBLrel. 31, Created)
DT   13-SEP-2005 (TrEMBLrel. 31, Last sequence update)
DT   13-SEP-2005 (TrEMBLrel. 31, Last annotation update)
DE   Hypothetical protein.
GN   ORFNames=UM01646.1;
OS   Ustilago maydis 521.
OC   Eukaryota; Fungi; Basidiomycota; Ustilaginomycetes;
OC   Ustilaginomycetidae; Ustilaginales; Ustilaginaceae; Ustilago.
OX   NCBI_TaxID=237631;
RN   [1]
RP   NUCLEOTIDE SEQUENCE.
RC   STRAIN=521;
RA   Birren B., Nusbaum C., Abebe A., Abouelleil A., Adekoya E.,
RA   Ait-zahra M., Allen N., Allen T., An P., Anderson M., Anderson S.,
RA   Arachchi H., Armbruster J., Bachantsang P., Baldwin J., Barry A.,
RA   Bayul T., Blitshsteyn B., Bloom T., Blye J., Boguslavskiy L.,
RA   Borowsky M., Boukhgalter B., Brunache A., Butler J., Calixte N.,
RA   Calvo S., Camarata J., Campo K., Chang J., Cheshatsang Y., Citroen M.,
RA   Collymore A., Considine T., Cook A., Cooke P., Corum B., Cuomo C.,
RA   David R., Dawoe T., Degray S., Dodge S., Dooley K., Dorje P.,
RA   Dorjee K., Dorris L., Duffey N., Dupes A., Elkins T., Engels R.,
RA   Erickson J., Farina A., Faro S., Ferreira P., Fischer H.,
RA   Fitzgerald M., Foley K., Gage D., Galagan J., Gearin G., Gnerre S.,
RA   Gnirke A., Goyette A., Graham J., Grandbois E., Gyaltsen K., Hafez N.,
RA   Hagopian D., Hagos B., Hall J., Hatcher B., Heller A., Higgins H.,
RA   Honan T., Horn A., Houde N., Hughes L., Hulme W., Husby E., Iliev I.,
RA   Jaffe D., Jones C., Kamal M., Kamat A., Kamvysselis M., Karlsson E.,
RA   Kells C., Kieu A., Kisner P., Kodira C., Kulbokas E., Labutti K.,
RA   Lama D., Landers T., Leger J., Levine S., Lewis D., Lewis T.,
RA   Lindblad-toh K., Liu X., Lokyitsang T., Lokyitsang Y., Lucien O.,
RA   Lui A., Ma L.J., Mabbitt R., Macdonald J., Maclean C., Major J.,
RA   Manning J., Marabella R., Maru K., Matthews C., Mauceli E.,
RA   Mccarthy M., Mcdonough S., Mcghee T., Meldrim J., Meneus L.,
RA   Mesirov J., Mihalev A., Mihova T., Mikkelsen T., Mlenga V., Moru K.,
RA   Mozes J., Mulrain L., Munson G., Naylor J., Newes C., Nguyen C.,
RA   Nguyen N., Nguyen T., Nicol R., Nielsen C., Nizzari M., Norbu C.,
RA   Norbu N., O'donnell P., Okoawo O., O'leary S., Omotosho B.,
RA   O'neill K., Osman S., Parker S., Perrin D., Phunkhang P., Piqani B.,
RA   Purcell S., Rachupka T., Ramasamy U., Rameau R., Ray V., Raymond C.,
RA   Retta R., Richardson S., Rise C., Rodriguez J., Rogers J., Rogov P.,
RA   Rutman M., Schupbach R., Seaman C., Settipalli S., Sharpe T.,
RA   Sheridan J., Sherpa N., Shi J., Smirnov S., Smith C., Sougnez C.,
RA   Spencer B., Stalker J., Stange-thomann N., Stavropoulos S.,
RA   Stetson K., Stone C., Stone S., Stubbs M., Talamas J., Tchuinga P.,
RA   Tenzing P., Tesfaye S., Theodore J., Thoulutsang Y., Topham K.,
RA   Towey S., Tsamla T., Tsomo N., Vallee D., Vassiliev H.,
RA   Venkataraman V., Vinson J., Vo A., Wade C., Wang S., Wangchuk T.,
RA   Wangdi T., Whittaker C., Wilkinson J., Wu Y., Wyman D., Yadav S.,
RA   Yang S., Yang X., Yeager S., Yee E., Young G., Zainoun J., Zembeck L.,
RA   Zimmer A., Zody M., Lander E.;
RT   "The genome sequence of Ustilago maydis.";
RL   Submitted (FEB-2004) to the EMBL/GenBank/DDBJ databases.
CC   -!- CAUTION: The sequence shown here is derived from an
CC       EMBL/GenBank/DDBJ whole genome shotgun (WGS) entry which is
CC       preliminary data.
DR   EMBL; AACP01000063; EAK82427.1; -; Genomic_DNA.
KW   Hypothetical protein.
SQ   SEQUENCE   1020 AA;  111539 MW;  86E5BB3D96E87470 CRC64;

  Query Match             41.0%;  Score 1890.5;  DB 2;  Length 1020;
  Best Local Similarity   42.2%;  Pred. No. 4.2e-96;
  Matches  425;  Conservative  169;  Mismatches  252;  Indels  161;  Gaps   23;

Qy         11 AVSSPYPSQSEGVSTPLPQVTSPSFDNAASPVAGRRAVRQTPTSAVRRRGRETDSARRRR 70

Db         51 ASSARAPQSSLGASSPLHFPTS-------SPSARTSRGQRTPLA----------SARRSR 93

Qy         71 SRSRSLGNSVYSSPYDAGTPGTPGTPVATPVYATPVGTPMGTPSFHRGTPQYKQRSELGS 130

Db         94 NAENRLPSS--SSAF------------------SRLDEPLFFPSSGGSTPRHQRRGEIHS 133

Qy        131 ----QGKPLHRRRRSQSREPGHRSPSREPSA--------DGRPSESAEPD---------- 168

Db        134 SIALSSPSLTRRNRADPSSQPQGAPTRSSSVFGGSQTAIDGHKASDAHSDALSFSQHGVT 193

Qy        169 -DTLG----GEYAYVWGTNVNIPDVLRAIRRFLHNYR----------------SSAHDLN 207

Db        194 SDAAGPGQDGVSKVIWGTNVSIGETMEMFRSFLRGFRLKYRWAHAKKLGEPLPSAATSNP 253

Qy        208 SKYIQIIEETVER-----EEDTLNIDMSDIYDHDPD--LYAKIVRYPLDIIPLLD----T 256

Db        254 AEGERLVYEGYLRRMRITDQTNLNLRISDLEAYPPSKRLKMQLIRYPQEMVPIMDQVLKD 313

Qy        257 ECQEVA------------------TSLLPTFEKHIEARPFNLKASVHMRELNPSDIDKL 297

Db        314 EMLEMAYEDQKEARDGMGGDMGLAEIELMET--KLYKVRPYGVEA-INMRELNPSDIDKL 370

Qy        298 VSVKGMVIRCSSIIPEIKGAFFKCLVCGHSPPLVTVVKGRVEEPTRCEKPECAARNAMSL 357

Db        371 VTVRGLVIRATPIIPEMKQAFFRCLVCNHTVP-VEIDRGRIAEPDRCPRQVCNLQGSMSL 429

Qy        358 IHNRCTFANKQIVRLQETPDAIPEGETPHTVSMCLYNTMVDAVKPGDRIEVTGVFKAMAV 417

Db        430 IHNRCEFSDRQVVRIQETPDVVPDGQTPHTVSMCAYDELVDVSKPGDRVEITGIFRSTPV 489

Qy        418 RVGPNQRTLRALYKTYIDCVHVKKSDRGR----LQTEDPMEM-------------DKEN 459

Db        490 RVNPRQRSLKSLYKTFVDILHIKRTNAKRLGVDLSTRDASEQAAGPGAQAVGVGGEEEDE 549

Qy        460 DMYAGYHESDTSEAANEAKIQ----KLKELSKLPGIYDRLSRSLAPSIWELEDIKKGLLC 515

Db        550 DVEVQSSHANDADDANVPRSQDLEDKLRSIAQRPDVYDVLSRSLAPSIYEMDDVKKGILL 609

Qy        516 QLFGGKAKKIPSGAS-----FRGDINVLLVGDPGTSKSQLLQYVHKIAPRGIYTSGRGSS 570

Db        610 QLFGGTNKTISTGGGGGGPRYRGDINVLMVGDPGIAKSQILQYVHKIAPRGVYASGKGSS 669

Qy        571 AVGLTAYVTKDPETRETVLESGALVLSDRGICCIDEFDKMSDNARSMLHEVMEQQTVSVA 630

Db        670 AVGLTAYVTRDPDTKQLVLESGALVLSDGGVCCIDEFDKMSEATRSVLHEVMEQQTLSIA 729

Qy        631 KGGIIASLNARTSVLACANPSGSRYNARLSVIDNIQLPPTLLSRFDLIYLMLDKPDEQND 690

Db        730 KAGIITTLNARASILAAANPTGSRYNVNLPITKNIDLPPTLISRFDLVYLVLDKIDEAND 789

Qy        691 RRLARHLVALHYENY-EVSKQDALDLQTLTAYITYARQHVHPTLSDEAAEDLINGYVEMR 749

Db        790 RRLARHLVSLYLEDKPDTGGKDVLPIETLTAYISYARNRLQPILTKEAGDALAARYVELR 849

Qy        750 QKGNFPGSSKKVITATPRQLESMIRISEALARMRFSEVVEKVDAAEAVRLLDVALQQSAT 809

Db        850 KVGEDPRNAERRITATTRQLESMIRLSEAHARMRFADEVIVDDVEEAARLIREAAKSSAT 909

Qy        810 DHATGTIDMDLITTGVSASERIRRANLLAALRELIADKISPGSSSGLKTS---------- 859

Db        910 DPRTGLIDLDLINTGRSYHQRKLAGDLRREFLQLLDEM---GSVSGRATSPSGPSSTQPT 966

Qy        860 ------QLLEDIRSQSSVDVSLQDIKNALGSLQGEGFLTVHGDIVKR 900

Db        967 KAVRYVDLAKALNDQSSVPVDQNDLHELVRTLESEGVIKTAGDRERR 1013


RESULT 5
Q4X1R4_ASPFU
```

```
ID   Q4X1R4_ASPFU PRELIMINARY;      PRT;  1023 AA.
AC   Q4X1R4;
DT   13-SEP-2005 (TrEMBLrel. 31, Created)
DT   13-SEP-2005 (TrEMBLrel. 31, Last sequence update)
DT   13-SEP-2005 (TrEMBLrel. 31, Last annotation update)
DE   DNA replication licensing factor Mcm4, putative.
GN   ORFNames=Afu2g09060;
OS   Aspergillus fumigatus Af293.
OC   Eukaryota; Fungi; Ascomycota; Pezizomycotina; Eurotiomycetes;
OC   Eurotiales; Trichocomaceae; mitosporic Trichocomaceae; Aspergillus.
OX   NCBI_TaxID=330879;
RN   [1]
RP   NUCLEOTIDE SEQUENCE.
RC   STRAIN=Af293;
RA   Nierman W., Pain A., Anderson M.J., Wortman J., Kim H.Stanley.,
RA   Arroya J., Berriman M., Abe K., Archer D.B., Bermejo C., Bennett J.,
RA   Bowyer P., Chen D., Collins M., Coulsen R., Davies R., Dyer P.S.,
RA   Farman M., Fedorova N., Fedorova N., Feldblyum T.V., Fischer R.,
RA   Fosker N., Fraser A., Garcia J.L., Garcia M.J., Goble A.,
RA   Goldman G.H., Gomi K., Griffith-Jones S., Gwilliam R., Haas B.,
RA   Haas H., Harris D., Horiuchi H., Huang J., Humphrey S., Jimenez J.,
RA   Keller N., Khouri H., Kitamoto K., Kobayashi T., Kulkarni R.,
RA   Kumagai T., Lafton A., Latge J.-P., Li W., Lord A., Lu C.,
RA   Majoros W.H., May G.S., Miller B.L., Mohamoud Y., Molina M., Monod M.,
RA   Mouyna I., Mulligan S., Murphy L., O'Neil S., Paulsen I.,
RA   Penalva M.A., Pertea M., Price C., Pritchard B.L., Quail M.A.,
RA   Rabbinowitsch E., Rawlins N., Rajandream M.-A., Reichard U.,
RA   Renauld H., Robson G.D., Rodriguez de Cordoba S., Rodriguez-Pena J.M.,
RA   Ronning C.M., Rutter S., Salzberg S.L., Sanchez M.,
RA   Sanchez-Ferrero J.C., Saunders D., Seeger K., Squares R., Squares S.,
RA   Takeuchi M., Tekaia F., Turner G., Vazquez de Aldana C.R., Weidman J.,
RA   White O., Woodward J., Yu J.-H., Fraser C., Galagan J.E., Asai K.,
RA   Machida M., Hall N., Barrell B., Denning D.W.;
RT   "Genomic sequence of the pathogenic and allergenic filamentous fungus
RT   Aspergillus fumigatus.";
RL   Submitted (MAY-2005) to the EMBL/GenBank/DDBJ databases.
CC   -!- CAUTION: The sequence shown here is derived from an
CC       EMBL/GenBank/DDBJ whole genome shotgun (WGS) entry which is
CC       preliminary data.
DR   EMBL; AAHF01000001; EAL93201.1; -; Genomic_DNA.
SQ   SEQUENCE   1023 AA;  113033 MW;  3397571B956DDFB6 CRC64;

  Query Match             40.9%;  Score 1884;  DB 2;  Length 1023;
  Best Local Similarity   42.4%;  Pred. No. 9.7e-96;
  Matches  435;  Conservative  154;  Mismatches  286;  Indels  152;  Gaps   23;

Qy         12 VSSPYPSQSEGVSTPLPQVTSPSFDNAASPVAGRRAVRQTPTSAVRRRGRETDSARRRRS 71
               :|||  |:  |                 |:  | | |       ||  ||:   |   :  |   |:
Db          1 MSSPASSRRRG---------RPAKDTATSSPARSTRSRQLQTSSPTPRAADEQSQATPRA 51

Qy         72 RSRSLGNSVY--SSPY-----------DAGTPGT----PGTPV----------ATPVYAT 104
                 |  |  :     |||             | ||       | :|:            ||
Db         52 SRRLRGEAAVPSSSPMFFQSSPSKADSSAETPDVRMDEPSSPMRESSTMDEGDRTPRGNA 111

Qy        105 PV---GTP---MGTPSFHRGTPQYKQRSELGSQGKPLH------------RRRRSQSREP 146
              |       :|    | : |   |     :   :||: : |     |                   |||
Db        112 PTMRDSSPIRYMSSSSPTRAQNRQSRRSDIPSSSSGLFVSSRPSIESNRAVSRRSDLHSG 171

Qy        147 GH-RSPSRE----------PSADGRPSESA-----EPD----DTLGGEYA-YVWGTNVNI 185
              |    ||:|            |: || |     |          ||       : |||        :||||::|
Db        172 GFLSSPNRRRRVFVDANGMPATDGDPRSDATFSNIHPDTSEAEALGGSSTRVIWGTNISI 231

Qy        186 PDVLRAIRRFLHN----YRSSAHDLNSKYIQIIEETVEREE-------------DTLNID 228
                | :  |  :  | ||:|     ||   |                :|: ::  |           :||:|
Db        232 QDSMSAFKNFLYNFQTKYRLWAEGATEDETRIMGDSAEEREYISMLSTMRQLGVTSLNLD 291

Qy        229 MSDI--YDHDPDLYAKIVRYPLDIIPLLDTECQEVATSL--------------------- 265
                ::    |      |: ::   || :||||:|      ::|      |
Db        292 AKNLKAYPSTLKLWHQLHAYPQEIIPLMDQTVKDVMVELAIKEMERLRAQNQRNQNHNRG 351

Qy        266 -----------------------LPTFEKHIEAR-----PFNLKASVHMRELNPSDIDKL 297
                                          :|      :|  :           || |  ::|:||:|:|:|:|||
Db        352 LSSGPAVPSSDALSETGRMPQNEIPDLVGEVETKAFKVLPFGLDSTVNMRDLDPADMDKL 411

Qy        298 VSVKGMVIRCSSIIPEIKGAFFKCLVCGHSPPLVTVVKGRVEEPTRCEKPECAARNAMSL 357
              ||:||:|||  :  |||:::|  |||:|  ||  |         |  :  :|:|  |||  |  :|  |     ||:|  |
Db        412 VSIKGLVIRTTPIIPDMKEAFFRCQVCNHGVQ-VDIDRGKVAEPTECPRPVCKERNSMQL 470

Qy        358 IHNRCTFANKQIVRLQETPDAIPEGETPHTVSMCLYNTMVDAVKPGDRIEVTGVFKAMAV 417
              |||||  ||:||:::||||||:||:|:|||:||:|:|:  :||    |  |||:||||:|:    |
Db        471 IHNRCVFADKQVIKLQETPDSIPDGQTPHSVSLCVYDELVDVCKAGDRVEVTGIFRCNPV 530

Qy        418 RVGPNQRTLRALYKTYIDCVHVKKSDRGRLQTEDPMEMDKENDMYAGYHESDTSEAANEA 477
              ||  |    |||   ::|:||||   :||:|  ||  |:|    :          :   ::     ||   |           |   |
Db        531 RVNPRQRTQKSLFKTYIDVLHVQKIDRKKLGIDVSTIEQELSEQAAGDAEQTRRLTAEEE 590

Qy        478 KIQKLKELSKLPGIYDRLSRSLAPSIWELEDIKKGLLCQLFGGKAKKIPSGAS--FRGDI 535
                :|:|    :      |  :|:  |||||||||:|||:|::|:|||:||   |||||    |       |   :  :||||
Db        591 --EKIKRTATRPDLYELLSRSLAPSIYEMDDVKKGILLQLFGGTNKTFQKGGNPRYRGDI 648

Qy        536 NVLLVGDPGTSKSQLLQYVHKIAPRGIYTSGRGSSAVGLTAYVTKDPETRETVLESGALV 595
              |:||  |||  ||||||:||||||||||:||||:||||||||||||||||:||||||: ||||||||
Db        649 NILLCGDPSTSKSQLLRYVHKIAPRGVYTSGKGSSAVGLTAYVTRDPETRQMVLESGALV 708

Qy        596 LSDRGICCIDEFDKMSDNARSMLHEVMEQQTVSVAKGGIIASLNARTSVLACANPSGSRY 655
              |||  ||||||||||:::  ||:|||||||||||:|| |||  :||||||:|| |||  ||||
Db        709 LSDGGICCIDEFDKMNESTRSVLHEVMEQQTVSIAKAGIITTLNARTSILASANPIGSRY 768

Qy        656 NARLSVIDNIQLPPTLLSRFDLIYLMLDKPDEQNDRRLARHLVALHYEN--YEVSKQDAL 713
              |   |   |   ||  |||||||||||:||:||:  |||  |||||:|||  ::  |:          ::|:  |
Db        769 NPNLPVPQNIDLPPTLLSRFDLVYLVLDRVDEQEDRRLAKHLVNMYLEDRPEHAAEQEIL 828

Qy        714 DLQTLTAYITYARQHVHPTLSDEAAEDLINGYVEMRQKGNFPGSSKKVITATPRQLESMI 773
                ::  ||||||||:   |||  |:    |  :   |    |  ||  |:  |:      ||  :   ||||  ||||||
Db        829 PIEFLTAYITYAKTKVHPVLTPAAGKALSDAYVNMRKLGDDIRSSDRRITATTRQLESMI 888

Qy        774 RISEALARMRFSEVVEKVDAAEAVRLLDVALQQSATDHATGTIDMDLITTGVSASERIRR 833
              |:|||  ||||  |    |        |     |||||:    |::|:|||   ||  |||  |:|  |  |||||   |
Db        889 RLSEAHARMRLSPEVTADDVEEAVRLIRSAIKQAATDSRTGLIDMSLLTEGTSASERRSR 948

Qy        834 ANLLAALRELIADKISPGSSSGLKTSQLLEDIRSQSSVDVSLQDIKNALGSLQGEGFLTV 893
                  |   ||   ::  |    |  |  ::    :  :::     :        ||::|         :|:  :|:  ||  ::|
Db        949 EALKRALLSVVDDLCSGGGAA--RWAEVFRILSENSSIEVDGAQFADAVRALEAEGAVSV 1006

Qy        894 HGDIVKR 900
               |:  :|
Db       1007 VGEGARR 1013


RESULT 6
Q7SHS5_NEUCR
ID   Q7SHS5_NEUCR PRELIMINARY;      PRT;  1013 AA.
AC   Q7SHS5;
DT   01-MAR-2004 (TrEMBLrel. 26, Created)
DT   01-MAR-2004 (TrEMBLrel. 26, Last sequence update)
DT   01-MAR-2004 (TrEMBLrel. 26, Last annotation update)
DE   Hypothetical protein.
GN   Name=NCU02539.1;
OS   Neurospora crassa.
OC   Eukaryota; Fungi; Ascomycota; Pezizomycotina; Sordariomycetes;
OC   Sordariomycetidae; Sordariales; Sordariaceae; Neurospora.
OX   NCBI_TaxID=5141;
RN   [1]
RP   NUCLEOTIDE SEQUENCE.
RC   STRAIN=OR74A;
RA   Galagan J.E., Calvo S.E., Borkovich K.A., Selker E.U., Read N.D.,
RA   Jaffe D., FitzHugh W., Ma L.-J., Smirnov S., Purcell S., Rehman B.,
RA   Elkins T., Engels R., Wang S., Nielsen C.B., Butler J., Endrizzi M.,
RA   Qui D., Ianakiev P., Pedersen D., Nelson M., Washburne M.,
RA   Selitrennikoff C.P., Kinsey J.A., Braun E.L., Zelter A., Schulte U.,
RA   Kothe G.O., Jedd G., Mewes W., Staben C., Marcotte E., Greenberg D.,
RA   Roy A., Foley K., Naylor J., Thomann N., Barrett R., Gnerre S.,
RA   Kamal M., Kamvysselis M., Mauceli E., Bielke C., Rudd S., Frishman D.,
RA   Krystofova S., Rasmussen C., Metzenberg R.L., Perkins D.D., Kroken S.,
RA   Cogoni C., Macino G., Catcheside D., Li W., Pratt R.J., Osmani S.A.,
RA   DeSouza C.C., Glass L., Orbach M.J., Berglund J., Voelker R.,
```

```
RA   Yarden O., Plamann M., Seiler S., Dunlap J., Radford A., Aramayo R.,
RA   Natvig D.O., Alex L.A., Mannhaupt G., Ebbole D.J., Freitag M.,
RA   Paulsen I., Sachs M.S., Lander E.S., Nusbaum C., Birren B.;
RT   "The Genome Sequence of the Filamentous Fungus Neurospora crassa.";
RL   Nature 0:0-0(2003).
CC   -!- CAUTION: The sequence shown here is derived from an
CC       EMBL/GenBank/DDBJ whole genome shotgun (WGS) entry which is
CC       preliminary data.
DR   EMBL; AABX01000002; EAA36434.1; -; Genomic_DNA.
DR   GO; GO:0005634; C:nucleus; IEA.
DR   GO; GO:0005524; F:ATP binding; IEA.
DR   GO; GO:0003677; F:DNA binding; IEA.
DR   GO; GO:0008094; F:DNA-dependent ATPase activity; IEA.
DR   GO; GO:0006270; P:DNA replication initiation; IEA.
DR   InterPro; IPR001208; MCM.
DR   InterPro; IPR008047; MCM_4.
DR   Pfam; PF00493; MCM; 1.
DR   PRINTS; PR01657; MCMFAMILY.
DR   PRINTS; PR01660; MCMPROTEIN4.
DR   ProDom; PD001041; MCM; 1.
DR   PROSITE; PS00847; MCM_1; 1.
DR   PROSITE; PS50051; MCM_2; 1.
KW   Hypothetical protein.
SQ   SEQUENCE   1013 AA;  111576 MW;  1980D7DB790AFF3E CRC64;

  Query Match             40.6%;  Score 1873;  DB 2;  Length 1013;
  Best Local Similarity   41.9%;  Pred. No. 3.9e-95;
  Matches  423;  Conservative  168;  Mismatches  257;  Indels  161;  Gaps   26;

Qy         18 SQSEGVSTPLPQVT-------SPSFDNAASPVAGRRAV----RQ------------TPTS 54
              |   | |:  |: |        || |   :||  |   |     ||            |:|
Db         23 SSQAGPSSATPRQTRASQLASSPLFYEPSSPANGAAPVSSPLRQMSNTQSTAHQGNAPSS 82

Qy         55 AVRRRGR---ETDSARRRRSRSRSLGNSVYSSPYDAGTPGTPGTPVATPVYATPVGTPMG 111
               :|::     : |   |   ||: :|:   |||    :  :||  :
Db         83 PLRQQTETQSDADRTPRANGRSQLIGD---SSPIRYASSSSPGRQL------------- 125

Qy        112 TPSFHRGTPQYKQRSE-----LGSQGKPLHRRRRSQSREPGHRSPSR------------ 153
                 |   |||      : ||      |||       |:|::
Db        126 -------TQQSDLRSESSQLFVSSQRSVAGRSRRGDINGDPLRTPAQIPRRIILDDAGRV 178

Qy        154 ---EPSADGRPSESAEPD----DTLGGE-YAYVWGTNVNIPDVLRAIRRFLHN----YRS 201
                 |  :|    :   |:     | |||:    ||||  :::  |    |  : ||  |     ||
Db        179 IRDAPGSDANSFVTNNPNTSEADALGGQSQGLVWGTTISLDDSFSAFKDFLRNFTRKYRM 238

Qy        202 SAH-----------DLNSK-YIQIIEETVEREEDTLNIDMSDI--YDHDPDLYAKIVRYP 247
              |           | :|| | :  :|  :  :  | |:|: |:  |        |: :   ||
Db        239 WADGADEAETIGHPDADSKPYWEALENMLLLGTNKLYLDLRDLKSYPRTLKLWHQAQHYP 298

Qy        248 LDIIPLLDT---EC------QEVAT------------------------------------ 263
               :|||::|      :|        :|:|:
Db        299 TEIIPVMDQCVHDCMMELAQKEMASQRASQNSRTAPGASQSSEPNFPSSERSEEPPTPRP 358

Qy        264 --SLLPTFEKHIE-----ARPFNLKASVHMRELNPSDIDKLVSVKGMVIRCSSIIPEIKG 316
                :   || |    :          ||: |        ::|:||||:||||:|||:||| :  :||::|
Db        359 AQTAAPTIEDQVSQMAYVVRPWGLDKITNLRDLNPSDMDKLVSIKGLVIRTTPVIPDMKD 418

Qy        317 AFFKCLVCGHSPPLVTVV--KGRVEEPTRCEKPECAARNAMSLIHNRCTFANKQIVRLQE 374
              ||||| |||||    :||   :|::  |||  |  :   ||::|:| :||||| | :||:::|||
Db        419 AFFKCSVCGHS---ITVQLDRGKIREPTECPRARCASKNSMQIIHNRCAFEDKQVIKLQE 475

Qy        375 TPDAIPEGETPHTVSMCLYNTMVDAVKPGDRIEVTGVFKAMAVRVGPNQRTLRALYKTYI 434
              ||| :| |:|||:||:|:|:|| :||   |||:|:|:||     |||  |   ||:::::|||:
Db        476 TPDNVPAGQTPHSVSVCVYNELVDFCKAGDRVELTGIFKVTPVRVNPRMRTVKSVHKTYV 535

Qy        435 DCVHVKKSDRGRLQTEDPMEMD--KENDMYAGYHESDTSEAANEAKIQKLKELSKLPGIY 492
              | |||:| || |: :  ||   :|  :|  : :|      |      :  : :::|| :  | ||
Db        536 DVVHVQKVDRKRMGS-DPSTLDLAEEEEAHANGQSMDEVRKVSPDEEERIKETAARPDIY 594

Qy        493 DRLSRSLAPSIWELEDIKKGLLCQLFGGKAKKIPSGAS--FRGDINVLLVGDPGTSKSQL 550
              | |||||||||:|::|:|||:| |||||   |       | |   :||||||| ||| ||||||
Db        595 DLLSRSLAPSIYEMDDVKKGILLQLFGGTNKTFEKGGSPKYRGDINVLLCGDPSTSKSQL 654

Qy        551 LQYVHKIAPRGIYTSGRGSSAVGLTAYVTKDPETRETVLESGALVLSDRGICCIDEFDKM 610
               | |||:|||||:||||:||||||||||||||:|||:|: ||||||||||  |:||||||||
Db        655 LSYVHRIAPRGVYTSGKGSSAVGLTAYVTRDPESRQLVLESGALVLSDGGVCCIDEFDKM 714

Qy        611 SDNARSMLHEVMEQQTVSVAKGGIIASLNARTSVLACANPSGSRYNARLSVIDNIQLPPT 670
              ::: ||:|||||||||||||| |||  :|||||:|||:| |||  |||||   |||   || ||||
Db        715 NESTRSVLHEVMEQQTVSVAKAGIITTLNARTSILASANPIGSRYNPDLSVPQNIDLPPT 774

Qy        671 LLSRFDLIYLMLDKPDEQNDRRLARHLVALHYENYEVSKQ---DALDLQTLTAYITYARQ 727
              ||||||:||:||: ||:||:|||||::::  |:    | |     |  |  ::  ||:||:|||
Db        775 LLSRFDLVYLILDRVDEKNDQRLARHLLSMYLEDKPESAQQANDVLPVEFLTSYISYARS 834

Qy        728 HVHPTLSDEAAEDLINGYVEMRQKGNFPGSSKKVITATPRQLESMIRISEALARMRFSEV 787
              |:|| |: ||  :|:: |||||: |     :::| ||||  ||||||||::|| |:|| |:
Db        835 HIHPALTPEAGRELVDAYVEMRKLGQDVRAAEKRITATTRQLESMIRLAEAHAKMRLSQT 894

Qy        788 VEKVDAAEAVRLLDVALQQSATDHATGTIDMDLITTGVSASERIRRANLLAALRELIADK 847
              | :  |  |||||:  ||: :|||  :  | |||  |:| |  ||:|| |:|::   |:  |: :
Db        895 VTRDDVREAVRLIKSALKTAATD-SQGRIDMSLLTEGTSAAERQRKADMKDAVIRLLDEM 953

Qy        848 ISPGSSSGLKTSQLLEDIRSQSSVDVSLQDIKNALGSLQGEGFLTVHGD 896
               | |   :: |::   :   : | |   :    : :|: || :  | |:
Db        954 TSGGQV--VRYSEVARRLGEGAGVQVEPAEFAEVMRALEMEGAVMVTGE 1000

RESULT 7
MCM4_SCHPO
ID   MCM4_SCHPO     STANDARD;      PRT;   931 AA.
AC   P29458; Q9P7K4; Q9USM0;
DT   01-APR-1993 (Rel. 25, Created)
DT   30-MAY-2000 (Rel. 39, Last sequence update)
DT   13-SEP-2005 (Rel. 48, Last annotation update)
DE   DNA replication licensing factor mcm4 (Minichromosome maintenance
DE   protein 4) (Cell division control protein 21).
GN   Name=mcm4; Synonyms=cdc21; ORFNames=SPCC16A11.17, SPCC24B10.01;
OS   Schizosaccharomyces pombe (Fission yeast).
OC   Eukaryota; Fungi; Ascomycota; Schizosaccharomycetes;
OC   Schizosaccharomycetales; Schizosaccharomycetaceae;
OC   Schizosaccharomyces.
OX   NCBI_TaxID=4896;
RN   [1]
RP   NUCLEOTIDE SEQUENCE.
RX   MEDLINE=93087163; PubMed=1454522;
RA   Coxon A., Maundrell K., Kearsey S.E.;
RT   "Fission yeast cdc21+ belongs to a family of proteins involved in an
RT   early step of chromosome replication.";
RL   Nucleic Acids Res. 20:5571-5577(1992).
RN   [2]
RP   SEQUENCE REVISION TO C-TERMINUS.
RA   Kearsey S.E.;
RL   Submitted (JAN-1999) to the EMBL/GenBank/DDBJ databases.
RN   [3]
RP   NUCLEOTIDE SEQUENCE [LARGE SCALE GENOMIC DNA].
RC   STRAIN=972;
RX   MEDLINE=21848401; PubMed=11859360; DOI=10.1038/nature724;
RA   Wood V., Gwilliam R., Rajandream M.A., Lyne M.H., Lyne R., Stewart A.,
RA   Sgouros J.G., Peat N., Hayles J., Baker S.G., Basham D., Bowman S.,
RA   Brooks K., Brown D., Brown S., Chillingworth T., Churcher C.M.,
RA   Collins M., Connor R., Cronin A., Davis P., Feltwell T., Fraser A.,
RA   Gentles S., Goble A., Hamlin N., Harris D.E., Hidalgo J., Hodgson G.,
RA   Holroyd S., Hornsby T., Howarth S., Huckle E.J., Hunt S., Jagels K.,
RA   James K.D., Jones L., Jones M., Leather S., McDonald S., McLean J.,
RA   Mooney P., Moule S., Mungall K.L., Murphy L.D., Niblett D., Odell C.,
RA   Oliver K., O'Neil S., Pearson D., Quail M.A., Rabbinowitsch E.,
RA   Rutherford K.M., Rutter S., Saunders D., Seeger K., Sharp S.,
RA   Skelton J., Simmonds M.N., Squares R., Squares S., Stevens K.,
RA   Taylor K., Taylor R.G., Tivey A., Walsh S.V., Warren T., Whitehead S.,
RA   Woodward J.R., Volckaert G., Aert R., Robben J., Grymonprez B.,
RA   Weltjens I., Vanstreels E., Rieger M., Schaefer M., Mueller-Auer S.,
RA   Gabel C., Fuchs M., Duesterhoeft A., Fritzc C., Holzer E., Moestl D.,
RA   Hilbert H., Borzym K., Langer I., Beck A., Lehrach H., Reinhardt R.,
RA   Pohl T.M., Eger P., Zimmermann W., Wedler H., Wambutt R., Purnelle B.,
RA   Goffeau A., Cadieu E., Dreano S., Gloux S., Lelaure V., Mottier S.,
```

```
RA   Galibert F., Aves S.J., Xiang Z., Hunt C., Moore K., Hurst S.M.,
RA   Lucas M., Rochet M., Gaillardin C., Tallada V.A., Garzon A., Thode G.,
RA   Daga R.R., Cruzado L., Jimenez J., Sanchez M., del Rey F., Benito J.,
RA   Dominguez A., Revuelta J.L., Moreno S., Armstrong J., Forsburg S.L.,
RA   Cerutti L., Lowe T., McCombie W.R., Paulsen I., Potashkin J.,
RA   Shpakovski G.V., Ussery D., Barrell B.G., Nurse P.;
RT   "The genome sequence of Schizosaccharomyces pombe.";
RL   Nature 415:871-880(2002).
RN   [4]
RP   SUBUNIT.
RC   STRAIN=SP011;
RX   MEDLINE=21518581; PubMed=11606526;
RA   Liang D.T., Forsburg S.L.;
RT   "Characterization of Schizosaccharomyces pombe mcm7(+) and cdc23(+)
RT   (MCM10) and interactions with replication checkpoints.";
RL   Genetics 159:471-486(2001).
CC   -!- FUNCTION: Required for S phase execution.
CC   -!- SUBUNIT: Heterohexamer. The heterodimers of mcm4/mcm6 and
CC       mcm3/mcm5 interact with mcm2 and mcm7.
CC   -!- SUBCELLULAR LOCATION: Nuclear.
CC   -!- SIMILARITY: Belongs to the MCM family.
CC   -!- SIMILARITY: Contains 1 MCM domain.
CC   --------------------------------------------------------------------------
CC   This Swiss-Prot entry is copyright. It is produced through a collaboration
CC   between  the Swiss Institute of Bioinformatics  and the  EMBL outstation -
CC   the European Bioinformatics Institute.  There are no  restrictions on  its
CC   use as long as its content is in no way modified and this statement is not
CC   removed.
CC   --------------------------------------------------------------------------
DR   EMBL; X58824; CAA41628.1; -; Genomic_DNA.
DR   EMBL; AL109957; CAB53089.1; -; Genomic_DNA.
DR   EMBL; AL157991; CAB76210.1; -; Genomic_DNA.
DR   PIR; S26640; S26640.
DR   GeneDB_Spombe; SPCC16A11.17; -.
DR   GO; GO:0042555; C:MCM complex; IDA.
DR   GO; GO:0006270; P:DNA replication initiation; IDA.
DR   InterPro; IPR001208; MCM.
DR   InterPro; IPR008047; MCM_4.
DR   InterPro; IPR012335; Thioredoxin_fold.
DR   Pfam; PF00493; MCM; 1.
DR   PRINTS; PR01657; MCMFAMILY.
DR   PRINTS; PR01660; MCMPROTEIN4.
DR   ProDom; PD001041; MCM; 1.
DR   SMART; SM00350; MCM; 1.
DR   PROSITE; PS00847; MCM_1; 1.
DR   PROSITE; PS50051; MCM_2; 1.
KW   ATP-binding; Complete proteome; DNA replication; DNA-binding;
KW   Nuclear protein; Nucleotide-binding; Transcription;
KW   Transcription regulation.
FT   DOMAIN      493    702       MCM.
FT   NP_BIND     545    552       ATP (Potential).
FT   CONFLICT    872    891       Missing (in Ref. 3).
SQ   SEQUENCE   931 AA;  103728 MW;  BCA9B045FC62811D CRC64;

  Query Match             40.6%;  Score 1870;  DB 1;  Length 931;
  Best Local Similarity   44.0%;  Pred. No. 5.1e-95;
  Matches  408;  Conservative  163;  Mismatches  238;  Indels  118;  Gaps   21;

Qy         38 AASPVAGRRAVRQTPTSAVRRRGRETDSARRRRSRSRSLGNSVYSSP--YDAGTPG---- 91
                ::|   :||      :||          ||    |:|         :| |||  : | :||
Db          2 SSSQQSGRANELRTP-------GRANSSSR----------EAVDSSPLFFPASSPGSTRL 44

Qy         92 -TPGTPVATPVYATPVGTPMGTP---------------------SFHRGTPQYKQRSEL 128
                || |    ||: ::|:                           | : ||:  :| ::
Db         45 TTPRTTARTPLASSPLLFESSSPGPNIPQSSRSHLLSQRNDLFLDSSSQRTPRSTRRGDI 104

Qy        129 GS--QGKPLHRRRRSQSREPGHRSPSR--EPSADGRPSESAEP-----DDTLGGEYAYVW 179
                |  |        |||    : ||  :||        :|      | |      |||:        :|
Db        105 HSSVQMSTPSRRREVDPQRPGVSTPSSLLFSGSDALTFSQAHPSSEVADDTV----RVIW 160

Qy        180 GTNVNIPDVLRAIRRFLHNYRS--------------SAHDLNSKYIQIIEETVEREEDTL 225
              ||||:| : : : | ||  ::                |  |   ||: :          : |
Db        161 GTNVSIQESIASFRGFLRGFKKKYRPEYRNELMPPPDAEQL--VYIEALRNMRIMGLEIL 218

Qy        226 NIDMSDIYDHDP--DLYAKIVRYPLDIIPLLDTECQEVATSLLPTFE----------KHI 273
              |:|: |:   :  |      || ::   || :|||::|     ::|    || |             |
Db        219 NLDVQDLKHYPPTKKLYHQLYSYPQEIIPIMDQTIKDVMLDLLGTNPPEDVLNDIELKIY 278

Qy        274 EARPFNLKASVHMRELNPSDIDKLVSVKGMVIRCSSIIPEIKGAFFKCLVCGHSPPLVTV 333
               : |||||:  ::||:||| |||||:|:||:|:||: :||::| |||:| ||||        | :
Db        279 KIRPFNLEKCINMRDLNPGDIDKLISIKGLVLRCTPVIPDMKQAFFRCSVCGHC-VTVEI 337

Qy        334 VKGRVEEPTRCEKPECAARNAMSLIHNRCTFANKQIVRLQETPDAIPEGETPHTVSMCLY 393
                :||: || :| :  | | ||| |||||  ||:||::||||| :|:|:|||:|:||:|:|
Db        338 DRGRIAEPIKCPREVCGATNAMQLIHNRSEFADKQVIKLQETPDVVPDGQTPHSVSLCVY 397

Qy        394 NTMVDAVKPGDRIEVTGVFKAMAVRVGPNQRTLRALYKTYIDCVHVKKSDRGRLQTEDPM 453
               : :||: : |||||||||:|: : ||: |   ||::|:|||:| ||:|| ||: ||  || ||
Db        398 DELVDSARAGDRIEVTGIFRCVPVRLNPRMRTVKSLFKTYVDVVHIKKQDKRRLGT-DPS 456

Qy        454 EMDKENDMYAGYHESDTSEAA----------NEAKIQKLKELSKLPGIYDRLSRSLAPSI 503
               :                  ||| :|| |          :: :::|::::||     ||| ||||||||
Db        457 TL-----------ESDIAEDAALQIDEVRKISDEEVEKIQQVSKRDDIYDILSRSLAPSI 505

Qy        504 WELEDIKKGLLCQLFGGKAKKIPSGAS--FRGDINVLLVGDPGTSKSQLLQYVHKIAPRG 561
               :|::|:||||| |||||   |       |||   :||||:|: ||| ||||||:|:||||||||
Db        506 YEMDDVKKGLLLQLFGGTNKSFHKGASPRYRGDINILMCGDPSTSKSQILKYVHKIAPRG 565

Qy        562 IYTSGRGSSAVGLTAYVTKDPETRETVLESGALVLSDRGICCIDEFDKMSDNARSMLHEV 621
               :||||:||||||||||:|:| :|:: |||||||||| |||||||||||||    ||:||||
Db        566 VYTSGKGSSAVGLTAYITRDQDTKQLVLESGALVLSDGGICCIDEFDKMSDATRSILHEV 625

Qy        622 MEQQTVSVAKGGIIASLNARTSVLACANPSGSRYNARLSVIDNIQLPPTLLSRFDLIYLM 681
               ||||||:|||| |||  :|||||:|| ||| |||  ||:|  |   |  |  ||||||||||:|:
Db        626 MEQQTVTVAKAGIITTLNARTSILASANPIGSKYNPDLPVTKNIDLPPTLLSRFDLVYLI 685

Qy        682 LDKPDEQNDRRLARHLVALHYEN--YEVSKQDALDLQTLTAYITYARQHVHPTLSDEAAE 739
               ||: ||  ||:|| |:|::: |:       :   :   :: ||:|||||| :::| :|:|||:
Db        686 LDRVDETLDRKLANHIVSMYMEDTPEHATDMEVFSVEFLTSYITYARNNINPVISEEAAK 745

Qy        740 DLINGYVEMRQKGNFPGSSKKVITATPRQLESMIRISEALARMRFSEVVEKVDAAEAVRL 799
               :|:| || ||: |    :|:| |||| ||||||||:||| |:|    |||  |  || ||
Db        746 ELVNAYVGMRKLGEDVRASEKRITATTRQLESMIRLSEAHAKMHLRNVVEVGDVLEAARL 805

Qy        800 LDVALQQSATDHATGTIDMDLITTGVSASERIRRANLLAALRELIADKISPGSSSGLKTS 859
               :  |:: ||| ||| | :|||  |:  | :   :::  |  ||::    | :  :  |
Db        806 IKTAIKDYATDPATGKISLDLIY--VNERETLVPEDMVKELANLISNLTVGGKT--MLVS 861

Qy        860 QLLEDIRSQSSVDVSLQDIKNALGSLQ 886
               |||  | |||   :    | :  ||: :
Db        862 QLLTRFREQSSTRLDASDFEACLGAFR 888


RESULT 8
Q55I48_CRYNE
ID   Q55I48_CRYNE PRELIMINARY;      PRT;   989 AA.
AC   Q55I48;
DT   13-SEP-2005 (TrEMBLrel. 31, Created)
DT   13-SEP-2005 (TrEMBLrel. 31, Last sequence update)
DT   13-SEP-2005 (TrEMBLrel. 31, Last annotation update)
DE   Hypothetical protein.
GN   ORFNames=CNBM1700;
OS   Cryptococcus neoformans var. neoformans B-3501A.
OC   Eukaryota; Fungi; Basidiomycota; Hymenomycetes; Heterobasidiomycetes;
OC   Tremellomycetidae; Tremellales; Tremellaceae; Filobasidiella.
OX   NCBI_TaxID=283643;
RN   [1]
RP   NUCLEOTIDE SEQUENCE.
RC   STRAIN=B-3501A;
RA   Fung E., Hyman R.W., Rowley D., Bruno D., Miranda M., Fukushima M.,
RA   Wickes B.L., Fu J., Davis R.W.;
RT   "Cryptococcus neoformans serotype D sequencing.";
RL   Submitted (JUL-2004) to the EMBL/GenBank/DDBJ databases.
CC   -!- CAUTION: The sequence shown here is derived from an
CC       EMBL/GenBank/DDBJ whole genome shotgun (WGS) entry which is
CC       preliminary data.
```

```
DR   EMBL; AAEY01000063; EAL17478.1; -; Genomic_DNA.
DR   GO; GO:0005634; C:nucleus; IEA.
DR   GO; GO:0005524; F:ATP binding; IEA.
DR   GO; GO:0003677; F:DNA binding; IEA.
DR   GO; GO:0008094; F:DNA-dependent ATPase activity; IEA.
DR   GO; GO:0006270; P:DNA replication initiation; IEA.
DR   InterPro; IPR001208; MCM.
DR   InterPro; IPR008047; MCM_4.
DR   Pfam; PF00493; MCM; 1.
DR   PRINTS; PR01657; MCMFAMILY.
DR   PRINTS; PR01660; MCMPROTEIN4.
DR   ProDom; PD001041; MCM; 1.
DR   SMART; SM00350; MCM; 1.
DR   PROSITE; PS00847; MCM_1; 1.
DR   PROSITE; PS50051; MCM_2; 1.
KW   Hypothetical protein.
SQ   SEQUENCE   989 AA;  108198 MW;  032EC9A65724D6F6 CRC64;

  Query Match            40.3%;  Score 1856.5;  DB 2;  Length 989;
  Best Local Similarity  41.5%;  Pred. No. 3.1e-94;
  Matches  427;  Conservative  157;  Mismatches  257;  Indels  187;  Gaps   24;

Qy         10 GAVSSP--YPSQSEGVSTPLPQV------------TSPSFDNAASPVAGRRAVR------ 49
               | |||  :|:  | |  :    :|              :||      :|| |  || |
Db          7 GRSSSPLHFPTSSVGGTPRASRVQQLSGSGAPAASSSPLHFPTSSPQASSRAARTPGGLR 66

Qy         50 -QTPTSAVRRRGRE--------TDSARRRRSRSRSLGNSV-YSSP--------------- 84
                 |:||  || |:             :  :   ||:|    : :|   :|||
Db         67 SDVPSSAAVRRIRDETPLFFPPSGGSTPRRARRGDIHSSFPHSSPSLARRQAQVPADSLS 126

Qy         85 YDAGTPGTPGTPVATPVYATPVGTPMGTPSFHRGTPQYKQRSELGSQGKPLHRRRRSQSR 144
                | :||                   ||| ||  |||:        ||     |    :||
Db        127 LRASSPG------------------MGTDSF--GTPRV-----FGSAAPTLSAVAQSQL- 160

Qy        145 EPGHRSPSREPSADGRPSESAEPDDTLGGEYAYVWGTNVNIPDVLRAIRRFL-------- 196
                           |||          | :  |    ::|||  ::: :  :    | ||
Db        161 -----------GADG---------DDIDGMVKFIWGTTISLQESMNLFRDFLRGFKPKYR 200

Qy        197 --HNYRSSAHDLNS---------KYIQIIEETVE-------------REEDTLNIDMSDI 232
                 :|   |  |    |           |  :  |  |               |  ||:|  ::
Db        201 AVYNAEQSRHAEESGGVAPPPMTLYDNLSAERAEVPLYETYLNRLRLTGETNLNLDALNL 260

Qy        233 YDHDP--DLYAKIVRYPLDIIPLLDTECQEV------------------------ATSLL 266
                 :  |     || ::| || ::||::|    ::|                          ||:
Db        261 LAYRPTKKLYQQLVNYPQEVIPIMDQVLRDVMIELGHEELEKAQTKFAEGNLSQLELSLI 320

Qy        267 PTFEKHIEA-----RPFNLKASVHMRELNPSDIDKLVSVKGMVIRCSSIIPEIKGAFFKC 321
                  : :|:         |||    : :|:|:||:||| |  ||||:|||:||| : :||::   |||:|
Db        321 TNEIRDVESRVYKVRPFGGEKTVNMRDLNPGDTDKLVTVKGLVIRATPVIPDMTTAFFRC 380

Qy        322 LVCGHSPPLVTVVKGRVEEPTRCEKPECAARNAMSLIHNRCTFANKQIVRLQETPDAIPE 381
               ||| |:       : :||: || ||  :      |  :        |||||||    | :||:||||||||:|:
Db        381 LVCQHTVQ-ADIDRGRISEPERCPRDVCGSTGTMSLIHNRSEFTDKQVIRLQETPDAVPD 439

Qy        382 GETPHTVSMCLYNTMVDAVKPGDRIEVTGVFKAMAVRVGPNQRTLRALYKTYIDCVHVKK 441
               |:||||||:|:|: :|| ||||||: :||:|:::  |||  |  ||::::|||||:| ||||:
Db        440 GQTPHTVSLCVYDELVDLVKPGDRVIITGIFRSIPVRVNPRQRSIKSLYKTYLDVVHVKR 499

Qy        442 SDRGRL-------------------QTEDPMEMDKENDMYAGYHE---SDTSEAANEAKI 479
               :: |:                    :|  |:    |  |   |   |    |   || |  :
Db        500 TNTARMGFDPSTRAGESKPPGVGVGGEDDEDELLARQDGDAAMDEDLGSPVRSAAAEME- 558

Qy        480 QKLKELSKLPGIYDRLSRSLAPSIWELEDIKKGLLCQLFGGKAKKIPSGAS-----FRGD 534
               |:: |||  | :|: |: ||||||:||||:||||:|||:| |||||  | |  |           :|||
Db        559 QRIIELSNHPDLYNILASSLAPSIYELEDVKKGILLQLFGGTNKSIARGGGGGPRYRGD 618

Qy        535 INVLLVGDPGTSKSQLLQYVHKIAPRGIYTSGRGSSAVGLTAYVTKDPETRETVLESGAL 594
               ||||:|||||||||||:|||||||||||:|||:|||||||||||:|||:::::  |||||||
Db        619 INVLMVGDPGTSKSQILQYVHKIAPRGVYTSGKGSSAVGLTAYVTRDPDSKQLVLESGAL 678

Qy        595 VLSDRGICCIDEFDKMSDNARSMLHEVMEQQTVSVAKGGIIASLNARTSVLACANPSGSR 654
               ||||  |:||||||||||  ||:|||||||||||:|| |||  :||||||:|| |||  ||
Db        679 VLSDGGVCCIDEFDKMSDATRSVLHEVMEQQTVSIAKAGIITTLNARTSILAAANPINSR 738

Qy        655 YNARLSVIDNIQLPPTLLSRFDLIYLMLDKPDEQNDRRLARHLVALHYENYEVSKQD-AL 713
               |:  |  :   || |||||:|||||:||:|||  || |||:||:|||| |:  :  |      |  :
Db        739 YDPNLPIPANIDLPPTLISRFDLLYLVLDKVDEVNDRKLAKHLVGLYLSDVEDQPADNII 798

Qy        714 DLQTLTAYITYARQHVHPTLSDEAAEDLINGYVEMRQKGNFPGSSKKVITATPRQLESMI 773
               |||||:||||||  :|| |:: |:| |:  |||||: |      : :| |||| |||||||
Db        799 PLQTLTSYITYARSKIHPVLTEGASEALVQAYVEMRKAGMDSRTQEKRITATTRQLESMI 858

Qy        774 RISEALARMRFSEVVEKVDAAEAVRLLDVALQQSATDHATGTIDMDLITTGVSASERIRR 833
               |: || |||| |: ||: |  |||||:  ||::||||  || ||:||| ||    :  |   |
Db        859 RLGEAHARMRLSDRVEEEDIREAVRLIKSALRESATDPLTGQIDLDLINTGAGQTMRRAR 918

Qy        834 ANLLAALRELIADKISPGSSSGLKTSQLLEDIRSQSSVDVSLQDIKNALGSLQGEGFLTV 893
               |:|    : :|: :|          | |:: : :::::  |||| |              :   |: |   : |
Db        919 ADLKREVIKLVVEK---ARSQGIRWAAVIDELNKQSSVPVDHAQFAEIVRELEEESIVKV 975

Qy        894 HGDIVKRV 901
               |:  :|:
Db        976 MGERERRI 983


RESULT 9
Q5K7N5_CRYNE
ID   Q5K7N5_CRYNE PRELIMINARY;      PRT;   989 AA.
AC   Q5K7N5;
DT   10-MAY-2005 (TrEMBLrel. 30, Created)
DT   10-MAY-2005 (TrEMBLrel. 30, Last sequence update)
DT   10-MAY-2005 (TrEMBLrel. 30, Last annotation update)
DE   DNA unwinding-related protein, putative.
GN   ORFNames=CNM01820;
OS   Cryptococcus neoformans var. neoformans JEC21.
OC   Eukaryota; Fungi; Basidiomycota; Hymenomycetes; Heterobasidiomycetes;
OC   Tremellomycetidae; Tremellales; Tremellaceae; Filobasidiella.
OX   NCBI_TaxID=214684;
RN   [1]
RP   NUCLEOTIDE SEQUENCE.
RC   STRAIN=JEC21;
RA   Loftus B., Amedeo P., Roncaglia P., Vamathevan J., Utterback T.,
RA   Van Aken S., Fraser C.;
RL   Submitted (MAY-2004) to the EMBL/GenBank/DDBJ databases.
RN   [2]
RP   NUCLEOTIDE SEQUENCE [LARGE SCALE GENOMIC DNA].
RC   STRAIN=JEC21;
RX   PubMed=15653466; DOI=10.1126/science.1103773;
RA   Loftus B.J., Fung E., Roncaglia P., Rowley D., Amedeo P., Bruno D.,
RA   Vamathevan J., Miranda M., Anderson I.J., Fraser J.A., Allen J.E.,
RA   Bosdet I.E., Brent M.R., Chiu R., Doering T.L., Donlin M.J.,
RA   D'Souza C.A., Fox D.S., Grinberg V., Fu J., Fukushima M., Haas B.J.,
RA   Huang J.C., Janbon G., Jones S.J.M., Koo H.L., Krzywinski M.I.,
RA   Kwon-Chung K.J., Lengeler K.B., Maiti R., Marra M.A., Marra R.E.,
RA   Mathewson C.A., Mitchell T.G., Pertea M., Riggs F.R., Salzberg S.L.,
RA   Schein J.E., Shvartsbeyn A., Shin H., Shumway M., Specht C.A.,
RA   Suh B.B., Tenney A., Utterback T.R., Wickes B.L., Wortman J.R.,
RA   Wye N.H., Kronstad J.W., Lodge J.K., Heitman J., Davis R.W.,
RA   Fraser C.M., Hyman R.W.;
RT   "The genome of the basidiomycetous yeast and human pathogen
RT   Cryptococcus neoformans.";
RL   Science 307:1321-1324(2005).
DR   EMBL; AE017353; AAW46787.1; -; Genomic_DNA.
DR   GO; GO:0005634; C:nucleus; IEA.
DR   GO; GO:0005524; F:ATP binding; IEA.
DR   GO; GO:0003677; F:DNA binding; IEA.
DR   GO; GO:0008094; F:DNA-dependent ATPase activity; IEA.
DR   GO; GO:0006270; P:DNA replication initiation; IEA.
DR   InterPro; IPR001208; MCM.
DR   InterPro; IPR008047; MCM_4.
DR   Pfam; PF00493; MCM; 1.
DR   PRINTS; PR01657; MCMFAMILY.
DR   PRINTS; PR01660; MCMPROTEIN4.
DR   ProDom; PD001041; MCM; 1.
DR   SMART; SM00350; MCM; 1.
DR   PROSITE; PS00847; MCM_1; 1.
```

```
DR   PROSITE; PS50051; MCM_2; 1.
KW   Complete proteome.
SQ   SEQUENCE   989 AA;  108198 MW;  032EC9A65724D6F6 CRC64;

  Query Match             40.3%;  Score 1856.5;  DB 2;  Length 989;
  Best Local Similarity   41.5%;  Pred. No. 3.1e-94;
  Matches  427;  Conservative  157;  Mismatches  257;  Indels  187;  Gaps   24;

Qy         10 GAVSSP--YPSQSEGVSTPLPQV------------TSPSFDNAASPVAGRRAVR------ 49
              | |||  :|: | | :     :|              :||     :|| |  || |
Db          7 GRSSSPLHFPTSSVGGTPRASRVQQLSGSGAPAASSSPLHFPTSSPQASSRAARTPGGLR 66

Qy         50 -QTPTSAVRRRGRE--------TDSARRRSRSRSLGNSV-YSSP--------------- 84
                 |:|| || |:          :  :  ||:|   : :|  :|||
Db         67 SDVPSSAAVRRIRDETPLFFPPSGGSTPRRARRGDIHSSFPHSSPSLARRQAQVPADSLS 126

Qy         85 YDAGTPGTPGTPVATPVYATPVGTPMGTPSFHRGTPQYKQRSELGSQGKPLHRRRRSQSR 144
                | :||                   ||| ||  |||:        ||       |    :||
Db        127 LRASSPG-------------------MGTDSF--GTPRV-----FGSAAPTLSAVAQSQL- 160

Qy        145 EPGHRSPSRREPSADGRPSESAEPDDTLGGEYAYVWGTNVNIPDVLRAIRRFL-------- 196
                          |||            | :  |     ::|||  ::: :  :   | ||
Db        161 -----------GADG---------DDIDGMVKFIWGTTISLQESMNLFRDFLRGFKPKYR 200

Qy        197 --HNYRSSAHDLNS---------KYIQIIEETVE-------------REEDTLNIDMSDI 232
                :|   |  |  |          |  :  |  |              |  ||:|  ::
Db        201 AVYNAEQSRHAEESGGVAPPPMTLYDNLSAERAEVPLYETYLNRLRLTGETNLNLDALNL 260

Qy        233 YDHDP--DLYAKIVRYPLDIIPLLDTECQEV------------------------ATSLL 266
                :  |   || ::| || ::||:|   ::|                           ||:
Db        261 LAYRPTKKLYQQLVNYPQEVIPIMDQVLRDVMIELGHEELEKAQTKFAEGNLSQLELSLI 320

Qy        267 PTFEKHIEA-----RPFNLKASVHMRELNPSDIDKLVSVKGMVIRCSSIIPEIKGAFFKC 321
                 : :|:       |||  : :|:|:||:||| | ||||:||:||| : :||::  |||:|
Db        321 TNEIRDVESRVYKVRPFGGEKTVNMRDLNPGDTDKLVTVKGLVIRATPVIPDMTTAFFRC 380

Qy        322 LVCGHSPPLVTVVKGRVEEPTRCEKPECAARNAMSLIHNRCTFANKQIVRLQETPDAIPE 381
              ||| |:   : :|: || ||  |   :   |||||| |  : |||||  :|||:||||||:|:
Db        381 LVCQHTVQ-ADIDRGRISEPERCPRDVCGSTGTMSLIHNRSEFTDKQVIRLQETPDAVPD 439

Qy        382 GETPHTVSMCLYNTMVDAVKPGDRIEVTGVFKAMAVRVGPNQRTLRALYKTYIDCVHVKK 441
              |:|||||:|:|: :|| ||||||: :||:|:::: ||| ||  |||:::::||||:| ||||:
Db        440 GQTPHTVSLCVYDELVDLVKPGDRVIITGIFRSIPVRVNPRQRSIKSLYKTYLDVVHVKR 499

Qy        442 SDRGRL-------------------QTEDPMEMDKENDMYAGYHE---SDTSEAANEAKI 479
              :: |:                      :|  |:    |   |   |   |    || |  :
Db        500 TNTARMGFDPSTRAGESKPPGVGVGGEDDEDELLARQDGDAAMDEDLGSPVRSAAAEME- 558

Qy        480 QKLKELSKLPGIYDRLSRSLAPSIWELEDIKKGLLCQLFGGKAKKIPSGAS-----FRGD 534
              |:: |||  | :|: |: |||||:||||:|||:|  |||||  | |  |           :|||
Db        559 QRIIELSNHPDLYNILASSLAPSIYELEDVKKGILLQLFGGTNKSIARGGGGGPRYRGD 618

Qy        535 INVLLVGDPGTSKSQLLQYVHKIAPRGIYTSGRGSSAVGLTAYVTKDPETRETVLESGAL 594
              ||||:|||||||||||:||||||||||:|||:|||||||||||||:|:||||:|::::|||||
Db        619 INVLMVGDPGTSKSQILQYVHKIAPRGVYTSGKGSSAVGLTAYVTRDPDSKQLVLESGAL 678

Qy        595 VLSDRGICCIDEFDKMSDNARSMLHEVMEQQTVSVAKGGIIASLNARTSVLACANPSGSR 654
              |||| |:|||||||||||  |||:||||||||||:|||||||:|||||  ||| |||
Db        679 VLSDGGVCCIDEFDKMSDATRSVLHEVMEQQTVSIAKAGIITTLNARTSILAAANPINSR 738

Qy        655 YNARLSVIDNIQLPPTLLSRFDLIYLMLDKPDEQNDRRLARHLVALHYENYEVSKQD-AL 713
              |:  | :  || |||||:||||||:|||:|||  || |||:||:||| |:  :  |     :
Db        739 YDPNLPIPANIDLPPTLISRFDLLYLVLDKVDEVNDRKLAKHLVGLYLSDVEDQPADNII 798

Qy        714 DLQTLTAYITYARQHVHPTLSDEAAEDLINGYVEMRQKGNFPGSSKKVITATPRQLESMI 773
               |||||:|||||  :|| |:: |:| |:  |||||: |     : :| |||| |||||||
Db        799 PLQTLTSYITYARSKIHPVLTEGASEALVQAYVEMRKAGMDSRTQEKRITATTRQLESMI 858

Qy        774 RISEALARMRFSEVVEKVDAAEAVRLLDVALQQSATDHATGTIDMDLITTGVSASERIRR 833
              |: || |||| |: ||:  |  |||||:  ||::||||  || ||:|||  ||   :  |  |
Db        859 RLGEAHARMRLSDRVEEEDIREAVRLIKSALRESATDPLTGQIDLDLINTGAGQTMRRAR 918

Qy        834 ANLLAALRELIADKISPGSSSGLKTSQLLEDIRSQSSVDVSLQDIKNALGSLQGEGFLTV 893
              |:|   : :|: :|       | |:: : :::::  ||||  |       :  |: | : |
Db        919 ADLKREVIKLVVEK---ARSQGIRWAAVIDELNKQSSVPVDHAQFAEIVRELEEESIVKV 975

Qy        894 HGDIVKRV 901
              |: :|:
Db        976 MGERERRI 983
```

```
RESULT 10
CDC54_YEAST
ID   CDC54_YEAST    STANDARD;      PRT;   933 AA.
AC   P30665;
DT   01-APR-1993 (Rel. 25, Created)
DT   01-NOV-1995 (Rel. 32, Last sequence update)
DT   13-SEP-2005 (Rel. 48, Last annotation update)
DE   Cell division control protein 54.
GN   Name=CDC54; Synonyms=HCD21; OrderedLocusNames=YPR019W;
GN   ORFNames=YP9531.13;
OS   Saccharomyces cerevisiae (Baker's yeast).
OC   Eukaryota; Fungi; Ascomycota; Saccharomycotina; Saccharomycetes;
OC   Saccharomycetales; Saccharomycetaceae; Saccharomyces.
OX   NCBI_TaxID=4932;
RN   [1]
RP   NUCLEOTIDE SEQUENCE.
RA   Dalton S.;
RL   Submitted (SEP-1994) to the EMBL/GenBank/DDBJ databases.
RN   [2]
RP   NUCLEOTIDE SEQUENCE [LARGE SCALE GENOMIC DNA].
RC   STRAIN=S288c / AB972;
RX   MEDLINE=97313271; PubMed=9169875;
RA   Bussey H., Storms R.K., Ahmed A., Albermann K., Allen E., Ansorge W.,
RA   Araujo R., Aparicio A., Barrell B.G., Badcock K., Benes V.,
RA   Botstein D., Bowman S., Brueckner M., Carpenter J., Cherry J.M.,
RA   Chung E., Churcher C.M., Coster F., Davis K., Davis R.W.,
RA   Dietrich F.S., Delius H., DiPaolo T., Dubois E., Duesterhoeft A.,
RA   Duncan M., Floeth M., Fortin N., Friesen J.D., Fritz C., Goffeau A.,
RA   Hall J., Hebling U., Heumann K., Hilbert H., Hillier L.W.,
RA   Hunicke-Smith S., Hyman R.W., Johnston M., Kalman S., Kleine K.,
RA   Komp C., Kurdi O., Lashkari D., Lew H., Lin A., Lin D., Louis E.J.,
RA   Marathe R., Messenguy F., Mewes H.-W., Mirtipati S., Moestl D.,
RA   Mueller-Auer S., Namath A., Nentwich U., Oefner P., Pearson D.,
RA   Petel F.X., Pohl T.M., Purnelle B., Rajandream M.A., Rechmann S.,
RA   Rieger M., Riles L., Roberts D., Schaefer M., Scharfe M., Scherens B.,
RA   Schramm S., Schroeder M., Sdicu A.-M., Tettelin H., Urrestarazu L.A.,
RA   Ushinsky S., Vierendeels F., Vissers S., Voss H., Walsh S.V.,
RA   Wambutt R., Wang Y., Wedler E., Wedler H., Winnett E., Zhong W.-W.,
RA   Zollner A., Vo D.H., Hani J.;
RT   "The nucleotide sequence of Saccharomyces cerevisiae chromosome XVI.";
RL   Nature 387:103-105(1997).
RN   [3]
RP   NUCLEOTIDE SEQUENCE OF 571-646.
RX   MEDLINE=93087163; PubMed=1454522;
RA   Coxon A., Maundrell K., Kearsey S.E.;
RT   "Fission yeast cdc21+ belongs to a family of proteins involved in an
RT   early step of chromosome replication.";
RL   Nucleic Acids Res. 20:5571-5577(1992).
CC   -!- FUNCTION: Required for S phase execution.
CC   -!- SUBCELLULAR LOCATION: Nuclear (By similarity).
CC   -!- SIMILARITY: Belongs to the MCM family.
CC   -!- SIMILARITY: Contains 1 MCM domain.
CC   --------------------------------------------------------------------------
CC   This Swiss-Prot entry is copyright. It is produced through a collaboration
CC   between  the Swiss Institute of Bioinformatics  and the  EMBL outstation -
CC   the European Bioinformatics Institute.  There are no  restrictions on  its
CC   use as long as its content is in no way modified and this statement is not
CC   removed.
CC   --------------------------------------------------------------------------
DR   EMBL; U14731; AAA86310.1; -; Genomic_DNA.
DR   EMBL; Z49919; CAA90164.1; -; Genomic_DNA.
DR   EMBL; Z71255; CAA95015.1; -; Genomic_DNA.
DR   EMBL; Z15032; CAA78750.1; -; Genomic_DNA.
DR   PIR; S56050; S56050.
DR   IntAct; P30665; -.
```

```
DR   GermOnline; 144284; -.
DR   Ensembl; YPR019W; Saccharomyces cerevisiae.
DR   SGD; S000006223; CDC54.
DR   GO; GO:0005737; C:cytoplasm; IDA.
DR   GO; GO:0042555; C:MCM complex; IDA.
DR   GO; GO:0005634; C:nucleus; IDA.
DR   GO; GO:0005656; C:pre-replicative complex; IDA.
DR   GO; GO:0004003; F:ATP-dependent DNA helicase activity; TAS.
DR   GO; GO:0003682; F:chromatin binding; TAS.
DR   GO; GO:0006270; P:DNA replication initiation; TAS.
DR   GO; GO:0006268; P:DNA unwinding; TAS.
DR   GO; GO:0006267; P:pre-replicative complex formation and maint. . .; IPI.
DR   InterPro; IPR001208; MCM.
DR   InterPro; IPR008047; MCM_4.
DR   Pfam; PF00493; MCM; 1.
DR   PRINTS; PR01657; MCMFAMILY.
DR   PRINTS; PR01660; MCMPROTEIN4.
DR   ProDom; PD001041; MCM; 1.
DR   SMART; SM00350; MCM; 1.
DR   PROSITE; PS00847; MCM_1; 1.
DR   PROSITE; PS50051; MCM_2; 1.
KW   ATP-binding; Cell cycle; Complete proteome; DNA replication;
KW   DNA-binding; Nuclear protein; Nucleotide-binding; Transcription;
KW   Transcription regulation.
FT   DOMAIN       518    725       MCM.
FT   NP_BIND      568    575       ATP (Potential).
SQ   SEQUENCE   933 AA;  105003 MW;  382D542AE38975E3 CRC64;

  Query Match             39.8%;  Score 1833;  DB 1;  Length 933;
  Best Local Similarity   40.9%;  Pred. No. 5.7e-93;
  Matches  393;  Conservative  178;  Mismatches  281;  Indels  110;  Gaps   14;

Qy          2 ENNDALDIGAVSSPYPSQSEGVSTPLPQVTSPSFDNAASPVAG----------------- 44
              |:|::       |||       : |     ||::||:   ::|    |
Db         11 EDNNS------SSPVVPNPDSVP---PQLSSPALFYSSSSSQGDIYGRNNSQNLSQGEGN 61

Qy         45 -RRAVRQTP---TSAVRRRGRETDSARRRRSRSRSLGNSVYSSPYDAGTPGTPGTPVATP 100
                | |:  :|     |: :|:  :     :: |: | ||   ::     | | :            | ::
Db         62 IRAAIGSSPLNFPSSSQRQNSDVFQSQGRQGRIRSSASASGRSRYHSDLRSDRALPTSS- 120

Qy        101 VYATPVGTPMGTPSFHRGTPQYKQRSELGSQGKPLHRRRRSQSREPGHRSPSREPADGR 160
                                                | ||  |:         ||:             ||      | |
Db        121 -------------------------SSLGRNGQNRVHMRRNDIHTSDLSSPRRIVDFDTR 155

Qy        161 --------------PSESAEPDDTLGGEYAYVWGTNVNIPDVLRAIRRFLHNYR------ 200
                                 |||::||              :|||||:|  :       | ||  :::
Db        156 SGVNTLDTSSSSAPPSEASEP-------LRIIWGTNVSIQECTTNFRNFLMSFKYKFRKI 208

Qy        201 --------SSAHDLNSKYIQIIEETVEREEDTLNIDMSDI--YDHDPDLYAKIVRYPLDI 250
                      ::  |      ||: :  |   |          ||:|   ::   |        ||| ::: || ::
Db        209 LDEREEFINNTTDEELYYIKQLNEMRELGTSNLNLDARNLLAYKQTEDLYHQLLNYPQEV 268

Qy        251 IPLLDTECQEVATSL---------LPTFE-KHIEARPFNLKASVHMRELNPSDIDKLVSV 300
              | ::|      ::      ||                  |       |  |   :  ||:|:  :      |||||:|||||:::
Db        269 ISIMDQTIKDCMVSLIVDNNLDYDLDEIETKFYKVRPYNVGSCKGMRELNPNDIDKLINL 328

Qy        301 KGMVIRCSSIIPEIKGAFFKCLVCGHSPPLVTVVKGRVEEPTRCEKPECAARNAMSLIHN 360
              ||:|:|  :  :||::|  |||||  ||  |:        |   :  :|  ::||  |||:  :|     |:||||||
Db        329 KGLVLRSTPVIPDMKVAFFKCNVCDHT-MAVEIDRGVIQEPARCERIDCNEPNSMSLIHN 387

Qy        361 RCTFANKQIVRLQETPDAIPEGETPHTVSMCLYNTMVDAVKPGDRIEVTGVFKAMAVRVG 420
              ||:||:||:::|||||  :|:|:|||::|:|:|:  :||:  :  ||||||| |:::  :|
Db        388 RCSFADKQVIKLQETPDFVPDGQTPHSISLCVYDELVDSCRAGDRIEVTGTFRSIPIRAN 447

Qy        421 PNQRTLRALYKTYIDCVHVKKSDRGRLQTEDPMEMDKENDMYAGYHESDTSEAANEAKIQ 480
                || |::||||:| |||||          ||    :            :            ::|   :              :   :
Db        448 SRQRVLKSLYKTYVDVVHVKKVSDKRLDVDTSTIEQELMQNKVDHNEVEEVRQITDQDLA 507

Qy        481 KLKELSKLPGIYDRLSRSLAPSIWELEDIKKGLLCQLFGGKAKKIPSGASFRGDINVLLV 540
              |::|::      :|   |:||:|||:||||:|||:|  |||||   |        |    :|||||:||
Db        508 KIREVAAREDLYSLLARSIAPSIYELEDVKKGILLQLFGGTNKTFTKGGRYRGDINILLC 567

Qy        541 GDPGTSKSQLLQYVHKIAPRGIYTSGRGSSAVGLTAYVTKDPETRETVLESGALVLSDRG 600
              ||| |||||:||||||| |||:||||:|||||||||:|:|  :|::  |||||||||| |
Db        568 GDPSTSKSQILQYVHKITPRGVYTSGKGSSAVGLTAYITRDVDTKQLVLESGALVLSDGG 627

Qy        601 ICCIDEFDKMSDNARSMLHEVMEQQTVSVAKGGIIASLNARTSVLACANPSGSRYNARLS 660
              :|||||||||||:  ||:|||||||||:|:|:|| ||| :||||:|:|| ||| |||||  |
Db        628 VCCIDEFDKMSDSTRSVLHEVMEQQTISIAKAGIITTLNARSSILASANPIGSRYNPNLP 687

Qy        661 VIDNIQLPPTLLSRFDLIYLMLDKPDEQNDRRLARHLVALHYEN--YEVSKQDALDLQTL 718
              | :|| ||| ||||||:||:|||  |||:||| ||:||  |: |:    :|: | | :: |
Db        688 VTENIDLPPPLLSRFDLVYLVLDKVDEKNDRELAKHLTNLYLEDKPEHISQDDVLPVEFL 747

Qy        719 TAYIITYARQHVHPTLSDEAAEDLINGYVEMRQKGNFPGSSKKVITATPRQLESMIRISEA 778
              | ||:||::|:|| :::  |    :|:    |  ||  |:      | :| ||||  |||||||:::|
Db        748 TMYISYAKEHIHPIITEAAKTELVRAYVGMRKMGDDSRSDEKRITATTRQLESMIRLAEA 807

Qy        779 LARMRFSEVVEKVDAAEAVRLLDVALQQSATDHATGTIDMDLITTGVSASERIRRANLLA 838
              |:|:    |||   |   |||||:   |::   |||   ||  |||:|: || |   :|   : :|
Db        808 HAKMKLKNVVELEDVQEAVRLIRSAIKDYATDPKTGKIDMNLVQTGKSVIQRKLQEDLSR 867

Qy        839 ALRELIADKISPGSSSGLKTSQLLEDIRSQSSVDVSLQDIKNALGSLQGEGFLTVHGDIV 898
               :  :: |:     :|  :  ::|::  |      |      |      ||: ||   || |    :  |  |: |
Db        868 EIMNVLKDQ----ASDSMSFNELIKQINEHSQDRVESSDIQEALSRLQQEDKVIVLGEGV 923

Qy        899 KR 900
               :|
Db        924 RR 925


RESULT 11
Q4IQK4_GIBZE
ID   Q4IQK4_GIBZE PRELIMINARY;      PRT;  1020 AA.
AC   Q4IQK4;
DT   13-SEP-2005 (TrEMBLrel. 31, Created)
DT   13-SEP-2005 (TrEMBLrel. 31, Last sequence update)
DT   13-SEP-2005 (TrEMBLrel. 31, Last annotation update)
DE   Hypothetical protein.
GN   ORFNames=FG00504.1;
OS   Gibberella zeae PH-1.
OC   Eukaryota; Fungi; Ascomycota; Pezizomycotina; Sordariomycetes;
OC   Hypocreomycetidae; Hypocreales; Nectriaceae; Gibberella.
OX   NCBI_TaxID=229533;
RN   [1]
RP   NUCLEOTIDE SEQUENCE.
RC   STRAIN=PH-1;
RA   Birren B., Nusbaum C., Abouelleil A., Allen N., Anderson S.,
RA   Arachchi H.M., Barna N., Bastien V., Bloom T., Boguslavkiy L.,
RA   Boukhgalter B., Butler J., Calvo S.E., Camarata J., Chang J.,
RA   Choepel Y., Collymore A., Cook A., Cooke P., Corum B., DeArellano K.,
RA   Diaz J.S., Dodge S., Dooley K., Dorris L., Elkins T., Engels R.,
RA   Erickson J., Faro S., Ferreira P., FitzGerald M., Gage D., Galagan J.,
RA   Gardyna S., Gnerre S., Graham L., Grand-Pierre N., Hafez N.,
RA   Hagopian D., Hagos B., Hall J., Horton L., Hulme W., Iliev I.,
RA   Jaffe D., Johnson R., Jones C., Kamal M., Kamat A., Karatas A.,
RA   Kells C., Landers T., Levine R., Lindblad-Toh K., Liu G., Lui A.,
RA   Ma L.-J., Mabbitt R., MacLean C., Macdonald P., Major J., Manning J.,
RA   Matthews C., Mauceli E., McCarthy M., Meldrim J., Meneus L.,
RA   Mihova T., Mlenga V., Murphy T., Naylor J., Nguyen C., Nicol R.,
RA   Nielsen C.B., Norbu C., O'Connor T., O'Donnell P., O'Neil D.,
RA   Oliver J., Peterson K., Phunkhang P., Pierre N., Purcell S.,
RA   Rachupka A., Ramasamy U., Raymond C., Retta R., Rise C., Rogov P.,
RA   Roman J., Schauer S., Schupback R., Seaman S., Severy P., Smirnov S.,
RA   Smith C., Spencer B., Stange-Thomann N., Stojanovic N., Stubbs M.,
RA   Talamas J., Tesfaye S., Theodore J., Topham K., Travers M.,
RA   Vassiliev H., Venkataraman V.S., Viel R., Vo A., Wang S., Wilson B.,
RA   Wu X., Wyman D., Young G., Zainoun J., Zembek L., Zimmer A., Zody M.,
RA   Lander E.;
RT   "Fusarium graminearum genome sequence.";
RL   Submitted (FEB-2004) to the EMBL/GenBank/DDBJ databases.
CC   -!- CAUTION: The sequence shown here is derived from an
CC       EMBL/GenBank/DDBJ whole genome shotgun (WGS) entry which is
CC       preliminary data.
DR   EMBL; AACM01000020; EAA68736.1; -; Genomic_DNA.
KW   Hypothetical protein.
```

```
SQ   SEQUENCE   1020 AA;  112421 MW;  C36A0F53C6037CD0 CRC64;

  Query Match             39.8%;  Score 1833;  DB 2;  Length 1020;
  Best Local Similarity   40.8%;  Pred. No. 6.5e-93;
  Matches  412;  Conservative  172;  Mismatches  285;  Indels  140;  Gaps   22;

Qy          4 NDALDIGAVSSPYPSQS----------EGVSTPLPQVTSPSFDNAASPVAGRRAVRQTPT 53
                |  |:: :    |||| |          |||:|  |: |      |       |  :  :|
Db         32 NSQLEVPSSPLPYPSSSPGPSGQNDLPRGVSSPFAQMMSSQSTRA------RSNIPSSPF 85

Qy         54 SAVRRRGRETDSARRRSRSRSLGNSVYSSPYDAGTPGTPG------TPVATPVYATPVG 107
                 :    :  |  |  |:    :         |||          :||       :  :  :      ||
Db         86 HQMTESQSDIDPQRTPRANPTTPAYHGGSSPVRYDPSSSPGRSLRHQSELRSEASGLFVG 145

Qy        108 TPMGTPSFHR---------GTPQYKQRSELGSQGKPLHRRRRSQSREPGHRSPSREPSAD 158
                : :|:|: ||              ||:  :|   |   |: :                 ||  ||
Db        146 SQIGSPAPHRRGDINSDAARTPRASRRVILDDVGRVV-----------------REGPAD 188

Qy        159 GRPSES---AEPD----DTLGGE-YAYVWGTNVNIPDVLRAIRRFLHN-------YRSSA 203
                |    |      :|:     | |||:  : :||| |:| |     : : || |           ||  |
Db        189 GSDVASFANRDPNTSEADVLGGQGQSLIWGTTVSIDDTFASFKDFLRNFTQKYRMYRDGA 248

Qy        204 H--------DLNSK-YIQIIEETVEREEDTLNIDMSDIYDHDP--DLYAKIVRYPLDIIP 252
                          |  || |  : :|   :         | :|:||:  :  |     |: :|    || :|:|
Db        249 SDEDVQNDPDAESKPYWEALENMLLLGTTRLYLDISDLNLYPPTRKLWHQIQAYPQEIVP 308

Qy        253 LLDTECQEVATSL-----------------------------LPTFEKHIEA-------- 275
               ::|    ::    :                                   |: ::   ||
Db        309 VMDQSVHDMMVEIARTETMRNRSQSTAGHQASQHSTQSSEPNFPSSDRPEEAPTPRAQPD 368

Qy        276 -----------------RPFNLKASVHMRELNPSDIDKLVSVKGMVIRCSSIIPEIKGAF 318
                                  |||  |    :  ::|:||||:|:|:|:|:||| :  :||:| ||
Db        369 QQQASLEDQVASSIYVLRPFGLDKTTNLRDLNPSDMDRLISIKGLVIRTTPVIPDMKDAF 428

Qy        319 FKCLVCGHSPPLVTVVKGRVEEPTRCEKPECAARNAMSLIHNRCTFANKQIVRLQETPDA 378
               |:| || ||        |  :  :|:: |||  |  :    |  :|:|  :|||||:  :||::::|||||||
Db        429 FRCNVCNHSVN-VGLDRGKIREPTECPREICKSKNSMLIIHNRCSFEDKQVIKLQETPDA 487

Qy        379 IPEGETPHTVSMCLYNTMVDAVKPGDRIEVTGVFKAMAVRVGPNQRTLRALYKTYIDCVH 438
               :|  |:|||:|||:|:||  :||    |  |||:|:||:|:    |||  |   ||  ::::|||:|  :|
Db        488 VPAGQTPHSVSVCVYNELVDFCKAGDRVELTGIFRVSPVRVNPAQRAVKSVHKTYVDVLH 547

Qy        439 VKKSDRGRLQTEDP--MEMDKENDMYAGYHESDTSEAANEAKIQKLKELSKLPGIYDRLS 496
               ::|  |: |:     ||    :  :    |  |   ||  :  :   :    :    :|::|  :       ||| ||
Db        548 IQKVDKRRMGA-DPSTLGIAGEEDAEAGENGIEETRKISIEDEEKIRETAARDDIYDLLS 606

Qy        497 RSLAPSIWELEDIKKGLLCQLFGGKAKKIPSGAS--FRGDINVLLVGDPGTSKSQLLQYV 554
               ||||||:|::|:|||:| |||||  |          |  |  :|||||||| ||| |:|||:| ||
Db        607 RSLAPSIYEMDDVKKGILLQLFGGTNKSFQKGGSPKYRGDINVLLCGDPSTAKSQMLSYV 666

Qy        555 HKIAPRGIYTSGRGSSAVGLTAYVTKDPETRETVLESGALVLSDRGICCIDEFDKMSDNA 614
               ||||||:||||:|||||||||||||:|||||: |||||||||  |:||||||||||
Db        667 HKIAPRGVYTSGKGSSAVGLTAYVTRDPETRQLVLESGALVLSDGGVCCIDEFDKMSDAT 726

Qy        615 RSMLHEVMEQQTVSVAKGGIIASLNARTSVLACANPSGSRYNARLSVIDNIQLPPTLLSR 674
               ||:|||||||||||||| |||  :||||||:||  |||  |||||   |    ||  |||||||
Db        727 RSVLHEVMEQQTVSVAKAGIITTLNARTSILASANPIGSRYNPDLPVPQNIDLPPTLLSR 786

Qy        675 FDLIYLMLDKPDEQNDRRLARHLVALHYENYEVS---KQDALDLQTLTAYITYARQHVHP 731
               |||:|||||  |||:||||||:||::|:   |:     |          |  |  ::  ||  ||:|||   :  |
Db        787 FDLVYLMLDTADEKNDRRLAKHLLSLYLEDKPQSAPTDNDILPVEFLTLYISYARSKIQP 846

Qy        732 TLSDEAAEDLINGYVEMRQKGNFPGSSKKVITATPRQLESMIRISEALARMRFSEVVEKV 791
               :| |||::|:    ||  ||     |     |: |  ||||  ||||||||::||  |:||   :|  ||  :
Db        847 VISQEAAQELVECYVAMRALGQDVRSADKRITATTRQLESMIRLAEAHAKMRLAETVTRD 906

Qy        792 DAAEAVRLLDVALQQSATDHATGTIDMDLITTGVSASERIRRANLLAALRELIADKISPG 851
                |  ||  ||:    ||: :|||  |  |  |||  |:|  |  ||::|  ||    |     |:    |: |:::  |
Db        907 DVREANRLIQSALKTAATD-ANGRIDMSLLTEGTSAADRKRREELRTAILHLL-DEMTAG 964

Qy        852 SSSGLKTSQLLEDIRSQSSVDVSLQDIKNALGSLQGEGFLTVHGDIVKR 900
               :: ::     :     :     :|:  |      :    :  :|:  |     |   :  |:    :|
Db        965 GNT-VRWGDVFRRLSEGASIPVEQSEFNEVMRALEAENALMIMGEGARR 1012
```

```
RESULT 12
MCM4_MOUSE
ID   MCM4_MOUSE      STANDARD;      PRT;   862 AA.
AC   P49717; O89056;
DT   01-OCT-1996 (Rel. 34, Created)
DT   01-OCT-1996 (Rel. 34, Last sequence update)
DT   13-SEP-2005 (Rel. 48, Last annotation update)
DE   DNA replication licensing factor MCM4 (CDC21 homolog) (P1-CDC21).
GN   Name=Mcm4; Synonyms=Cdc21, Mcmd4;
OS   Mus musculus (Mouse).
OC   Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC   Mammalia; Eutheria; Euarchontoglires; Glires; Rodentia; Sciurognathi;
OC   Muroidea; Muridae; Murinae; Mus.
OX   NCBI_TaxID=10090;
RN   [1]
RP   NUCLEOTIDE SEQUENCE.
RX   MEDLINE=95334361; PubMed=7610039;
RA   Kimura H., Takizawa N., Nozaki N., Sugimoto K.;
RT   "Molecular cloning of cDNA encoding mouse Cdc21 and CDC46 homologs and
RT   characterization of the products: physical interaction between
RT   P1(MCM3) and CDC46 proteins.";
RL   Nucleic Acids Res. 23:2097-2104(1995).
RN   [2]
RP   NUCLEOTIDE SEQUENCE OF 503-602.
RC   STRAIN=BALB/c; TISSUE=Spleen;
RX   MEDLINE=99012997; PubMed=9798653; DOI=10.1016/S0161-5890(98)00031-5;
RA   Chu C.C., Paul W.E.;
RT   "Expressed genes in interleukin-4 treated B cells identified by cDNA
RT   representational difference analysis.";
RL   Mol. Immunol. 35:487-502(1998).
CC   -!- FUNCTION: Involved in the control of DNA replication.
CC   -!- SUBCELLULAR LOCATION: Nuclear (By similarity).
CC   -!- SIMILARITY: Belongs to the MCM family.
CC   -!- SIMILARITY: Contains 1 MCM domain.
CC   --------------------------------------------------------------------------
CC   This Swiss-Prot entry is copyright. It is produced through a collaboration
CC   between  the Swiss Institute of Bioinformatics  and the  EMBL outstation -
CC   the European Bioinformatics Institute.  There are no  restrictions on  its
CC   use as long as its content is in no way modified and this statement is not
CC   removed.
CC   --------------------------------------------------------------------------
DR   EMBL; D26089; BAA05082.1; -; mRNA.
DR   EMBL; U89402; AAC36509.1; -; mRNA.
DR   PIR; S56766; S56766.
DR   Ensembl; ENSMUSG00000022673; Mus musculus.
DR   MGI; MGI:103199; Mcm4.
DR   InterPro; IPR003593; AAA_ATPase.
DR   InterPro; IPR001208; MCM.
DR   InterPro; IPR008047; MCM_4.
DR   Pfam; PF00493; MCM; 1.
DR   PRINTS; PR01657; MCMFAMILY.
DR   PRINTS; PR01660; MCMPROTEIN4.
DR   ProDom; PD001041; MCM; 1.
DR   SMART; SM00382; AAA; 1.
DR   SMART; SM00350; MCM; 1.
DR   PROSITE; PS00847; MCM_1; 1.
DR   PROSITE; PS50051; MCM_2; 1.
KW   ATP-binding; DNA replication; DNA-binding; Nuclear protein;
KW   Nucleotide-binding; Transcription; Transcription regulation.
FT   DOMAIN      457    666       MCM.
FT   NP_BIND     509    516       ATP (Potential).
FT   CONFLICT    530    530       Q -> R (in Ref. 2).
FT   CONFLICT    572    572       I -> T (in Ref. 2).
SQ   SEQUENCE   862 AA;  96736 MW;  516ACC1A3C6FB16E CRC64;

  Query Match             39.5%;  Score 1821;  DB 1;  Length 862;
  Best Local Similarity   43.7%;  Pred. No. 2.4e-92;
  Matches  401;  Conservative  154;  Mismatches  258;  Indels  104;  Gaps   18;

Qy         30 VTSPSFDNAASPVAGRRAVRQ---TPTSAVRRRGRETDSARRRSRSRSLG--------- 77
                 ::||   |::|   ||: |:    ||| ::|      :    ||||     | |
```

```
Db          1 MSSP----ASTP--SRRSSRRGRVTPTQSLRSEESRSSPNRRRRGEDSSTGELLPMPTSP 54

Qy         78 ----------NSVYSSP-----------YDAGTPGTPGTPVATPVYATPVGTPMGTP-SF 115
                        |:::|||            :|   :| |         | ||     ||| |
Db         55 GADLQSPPAQNALFSSPPQMHSLAIPLDFDVSSPLT---------YGTPSSRVEGTPRSG 105

Qy        116 HRGTPQYKQRSELGSQGKPLHRRRRSQSREPGHRSPSREPSADGRPSESAEPDDTLGGEY 175
              ||||  :|| :|||  | |      :|                ||  :|    | :   |:
Db        106 VRGTP-VRQRPDLGSARKGLQVDLQS----------------DGAAAEDIVPSEQSLGQK 148

Qy        176 AYVWGTNVNIPDVLRAIRRFLHNY--------RSSAHDLNSK-YIQIIEETVEREEDTLN 226
               :|||:||:          :|||  :          :    |:   |:| : |    |  ||
Db        149 LVIWGTDVNVATCKENFQRFLQCFTDPLAKEEENVGIDITQPLYMQQLGEINITGEPFLN 208

Qy        227 IDMSDIYDHDPDLYAKIVRYPLDIIPLLDTECQEVATSLLP--TFEKHIEARPFNLKASV 284
              ::    |       :|| :::  || ::||    |         |:          |      |    |  |: ||||       :
Db        209 VNCEHIKSFSKNLYRQLISYPQEVIPTFDMAVNEIFFDRYPDSILEHQIQVRPFNALKTK 268

Qy        285 HMRELNPSDIDKLVSVKGMVIRCSSIIPEIKGAFFKCLVCGHSPPLVTVVKGRVEEPTRC 344
               || ||| |||:|::: ||||| | :|||:: |||:| || |:    | : :||: || |
Db        269 SMRNLNPEDIDQLITISGMVIRTSQLIPEMQEAFFQCQVCAHT-TRVEIDRGRIAEPCSC 327

Qy        345 EKPECAARNAMSLIHNRCTFANKQIVRLQETPDAIPEGETPHTVSMCLYNTMVDAVKPGD 404
                   |     ::|:|||||  |::||::|||:|: :| |:||||: :  :| :|| |:|||
Db        328 --VHCHTTHSMALIHNRSFFSDKQMIKLQESPEDMPAGQTPHTIVLFAHNDLVDKVQPGD 385

Qy        405 RIEVTGVFKAMAVRVGPNQRTLRALYKTYIDCVHVKKSDRGRLQTEDPMEMDKENDMYAG 464
              |: |||:::|: :|| |     ::::|||:||  :| :|:|  ||                |
Db        386 RVNVTGIYRAVPIRVNPRVSNVKSVYKTHIDVIHYRKTDAKRLH---------------G 430

Qy        465 YHESDTSEAANEAKIQKLKELSKLPGIYDRLSRSLAPSIWELEDIKKGLLCQLFGGKAKK 524
                    |     :   | :::  |||||:   ||:||:  :|||||:| ||||||:| |||||   |
Db        431 LDEEAEQKLFSEKRVKLLKELSRKPDIYERLASALAPSIYEHEDIKKGILLQLFGGTRKD 490

Qy        525 I--PSGASFRGDINVLLVGDPGTSKSQLLQYVHKIAPRGIYTSGRGSSAVGLTAYVTKDP 582
                        || :||:||  |||||||||||||: :  ||| ||||:||||||||||| |||
Db        491 FSHTGRGKFRAEINILLCGDPGTSKSQLLQYVYNLVPRGQYTSGKGSSAVGLTAYVMKDP 550

Qy        583 ETRETVLESGALVLSDRGICCIDEFDKMSDNARSMLHEVMEQQTVSVAKGGIIASLNART 642
              |||: ||::||||||| ||||||||||:::  ||:||||||||:|:|| |||  |||||
Db        551 ETRQLVLQTGALVLSDNGICCIDEFDKMNESTRSVLHEVMEQQTLSIAKAGIICQLNART 610

Qy        643 SVLACANPSGSRYNARLSVIDNIQLPPTLLSRFDLIYLMLDKPDEQNDRRLARHLVALHY 702
              |||| |||   |::| : : |:||||| ||||||||||:||||   ||  |||||  |||:|:|
Db        611 SVLAAANPIESQWNPKKTTIENIQLPHTLLSRFDLIFLMLDPQDEAYDRRLAHHLVSLYY 670

Qy        703 ENYEVSKQDALDLQTLTAYITYARQHVHPTLSDEAAEDLINGYVEMRQKGNFPGSSKKVI 762
              :: |   :::  ||:    |    || ||    | ||:|  ||:         |||:  ::
Db        671 QSEEQVEEEFLDMAVLKDYIAYAHSTIMPRLSEEASQALIEAYVNMRK----IGSSRGMV 726

Qy        763 TATPRQLESMIRISEALARMRFSEVVEKVDAAEAVRLLDVALQQSATDHATGTIDMDLIT 822
              :| ||||||:||:::|| |::|||  ||  :|    || ||   ||:|||||   ||  :|: ::|
Db        727 SAYPRQLESLIRLAEAHAKVRFSNKVEAIDVEEAKRLHREALKQSATDPRTGIVDISILT 786

Qy        823 TGVSASERIRRANLLAALRELIADKISPGSSSGLKTSQLLEDIRSQSSVDVSLQDIKNAL 882
              ||:||: | |:  |  |||:||  :| | :   || || |||| ||    ::     : ||
Db        787 TGMSATSRKRKEELAEALRKLI---LSKGKTPALKYQQLFEDIRGQSDTAITKDMFEEAL 843

Qy        883 GSLQGEGFLTVHGDIVK 899
               :|  : |||| |  |:
Db        844 RALADDDFLTVTGKTVR 860
```

RESULT 13
Q542F4_MOUSE
ID   Q542F4_MOUSE PRELIMINARY;   PRT;   862 AA.
AC   Q542F4;
DT   13-SEP-2005 (TrEMBLrel. 31, Created)
DT   13-SEP-2005 (TrEMBLrel. 31, Last sequence update)
DT   13-SEP-2005 (TrEMBLrel. 31, Last annotation update)
DE   2 days neonate thymus thymic cells cDNA, RIKEN full-length enriched
DE   library, clone:E430026B21 product:mini chromosome maintenance
DE   deficient 4 homolog (S. cerevisiae), full insert sequence.
GN   Name=Mcm4;
OS   Mus musculus (Mouse).
OC   Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC   Mammalia; Eutheria; Euarchontoglires; Glires; Rodentia; Sciurognathi;
OC   Muridae; Murinae; Mus.
OX   NCBI_TaxID=10090;
RN   [1]
RP   NUCLEOTIDE SEQUENCE.
RC   STRAIN=NOD; TISSUE=Thymus;
RX   MEDLINE=99279253; PubMed=10349636; DOI=10.1016/S0076-6879(99)03004-9;
RA   Carninci P., Hayashizaki Y.;
RT   "High-efficiency full-length cDNA cloning.";
RL   Meth. Enzymol. 303:19-44(1999).
RN   [2]
RP   NUCLEOTIDE SEQUENCE.
RC   STRAIN=NOD; TISSUE=Thymus;
RX   MEDLINE=21085660; PubMed=11217851; DOI=10.1038/35055500;
RA   Kawai J., Shinagawa A., Shibata K., Yoshino M., Itoh M., Ishii Y.,
RA   Arakawa T., Hara A., Fukunishi Y., Konno H., Adachi J., Fukuda S.,
RA   Aizawa K., Izawa M., Nishi K., Kiyosawa H., Kondo S., Yamanaka I.,
RA   Saito T., Okazaki Y., Gojobori T., Bono H., Kasukawa T., Saito R.,
RA   Kadota K., Matsuda H.A., Ashburner M., Batalov S., Casavant T.,
RA   Fleischmann W., Gaasterland T., Gissi C., King B., Kochiwa H.,
RA   Kuehl P., Lewis S., Matsuo Y., Nikaido I., Pesole G., Quackenbush J.,
RA   Schriml L.M., Staubli F., Suzuki R., Tomita M., Wagner L., Washio T.,
RA   Sakai K., Okido T., Furuno M., Aono H., Baldarelli R., Barsh G.,
RA   Blake J., Boffelli D., Bojunga N., Carninci P., de Bonaldo M.F.,
RA   Brownstein M.J., Bult C., Fletcher C., Fujita M., Gariboldi M.,
RA   Gustincich S., Hill D., Hofmann M., Hume D.A., Kamiya M., Lee N.H.,
RA   Lyons P., Marchionni L., Mashima J., Mazzarelli J., Mombaerts P.,
RA   Nordone P., Ring B., Ringwald M., Rodriguez I., Sakamoto N.,
RA   Sasaki H., Sato K., Schoenbach C., Seya T., Shibata Y., Storch K.-F.,
RA   Suzuki H., Toyo-oka K., Wang K.H., Weitz C., Whittaker C., Wilming L.,
RA   Wynshaw-Boris A., Yoshida K., Hasegawa Y., Kawaji H., Kohtsuki S.,
RA   Hayashizaki Y.;
RT   "Functional annotation of a full-length mouse cDNA collection.";
RL   Nature 409:685-690(2001).
RN   [3]
RP   NUCLEOTIDE SEQUENCE.
RC   STRAIN=NOD; TISSUE=Thymus;
RX   MEDLINE=22354683; PubMed=12466851; DOI=10.1038/nature01266;
RA   Okazaki Y., Furuno M., Kasukawa T., Adachi J., Bono H., Kondo S.,
RA   Nikaido I., Osato N., Saito R., Suzuki H., Yamanaka I., Kiyosawa H.,
RA   Yagi K., Tomaru Y., Hasegawa Y., Nogami A., Schonbach C., Gojobori T.,
RA   Baldarelli R., Hill D.P., Bult C., Hume D.A., Quackenbush J.,
RA   Schriml L.M., Kanapin A., Matsuda H., Batalov S., Beisel K.W.,
RA   Blake J.A., Bradt D., Brusic V., Chothia C., Corbani L.E., Cousins S.,
RA   Dalla E., Dragani T.A., Fletcher C.F., Forrest A., Frazer K.S.,
RA   Gaasterland T., Gariboldi M., Gissi C., Godzik A., Gough J.,
RA   Grimmond S., Gustincich S., Hirokawa N., Jackson I.J., Jarvis E.D.,
RA   Kanai A., Kawaji H., Kawasawa Y., Kedzierski R.M., King B.L.,
RA   Konagaya A., Kurochkin I.V., Lee Y., Lenhard B., Lyons P.A.,
RA   Maglott D.R., Maltais L., Marchionni L., McKenzie L., Miki H.,
RA   Nagashima T., Numata K., Okido T., Pavan W.J., Pertea G., Pesole G.,
RA   Petrovsky N., Pillai R., Pontius J.U., Qi D., Ramachandran S.,
RA   Ravasi T., Reed J.C., Reed D.J., Reid J., Ring B.Z., Ringwald M.,
RA   Sandelin A., Schneider C., Semple C.A., Setou M., Shimada K.,
RA   Sultana R., Takenaka Y., Taylor M.S., Teasdale R.D., Tomita M.,
RA   Verardo R., Wagner L., Wahlestedt C., Wang Y., Watanabe Y., Wells C.,
RA   Wilming L.G., Wynshaw-Boris A., Yanagisawa M., Yang I., Yang L.,
RA   Yuan Z., Zavolan M., Zhu Y., Zimmer A., Carninci P., Hayatsu N.,
RA   Hirozane-Kishikawa T., Konno H., Nakamura M., Sakazume N., Sato K.,
RA   Shiraki T., Waki K., Kawai J., Aizawa K., Arakawa T., Fukuda S.,
RA   Hara A., Hashizume W., Imotani K., Ishii Y., Itoh M., Kagawa I.,
RA   Miyazaki A., Sakai K., Sasaki D., Shibata K., Shinagawa A.,
RA   Yasunishi A., Yoshino M., Waterston R., Lander E.S., Rogers J.,
RA   Birney E., Hayashizaki Y.;
RT   "Analysis of the mouse transcriptome based on functional annotation of
RT   60,770 full-length cDNAs.";
RL   Nature 420:563-573(2002).
RN   [4]
RP   NUCLEOTIDE SEQUENCE.
RC   STRAIN=NOD; TISSUE=Thymus;

```
RX   MEDLINE=20499374; PubMed=11042159; DOI=10.1101/gr.145100;
RA   Carninci P., Shibata Y., Hayatsu N., Sugahara Y., Shibata K., Itoh M.,
RA   Konno H., Okazaki Y., Muramatsu M., Hayashizaki Y.;
RT   "Normalization and subtraction of cap-trapper-selected cDNAs to
RT   prepare full-length cDNA libraries for rapid discovery of new genes.";
RL   Genome Res. 10:1617-1630(2000).
RN   [5]
RP   NUCLEOTIDE SEQUENCE.
RC   STRAIN=NOD; TISSUE=Thymus;
RX   MEDLINE=20530913; PubMed=11076861; DOI=10.1101/gr.152600;
RA   Shibata K., Itoh M., Aizawa K., Nagaoka S., Sasaki N., Carninci P.,
RA   Konno H., Akiyama J., Nishi K., Kitsunai T., Tashiro H., Itoh M.,
RA   Sumi N., Ishii Y., Nakamura S., Hazama M., Nishine T., Harada A.,
RA   Yamamoto R., Matsumoto H., Sakaguchi S., Ikegami T., Kashiwagi K.,
RA   Fujiwake S., Inoue K., Togawa Y., Izawa M., Ohara E., Watahiki M.,
RA   Yoneda Y., Ishikawa T., Ozawa K., Tanaka T., Matsuura S., Kawai J.,
RA   Okazaki Y., Muramatsu M., Inoue Y., Kira A., Hayashizaki Y.;
RT   "RIKEN integrated sequence analysis (RISA) system-384-format
RT   sequencing pipeline with 384 multicapillary sequencer.";
RL   Genome Res. 10:1757-1771(2000).
RN   [6]
RP   NUCLEOTIDE SEQUENCE.
RC   STRAIN=NOD; TISSUE=Thymus;
RA   Adachi J., Aizawa K., Akimura T., Arakawa T., Bono H., Carninci P.,
RA   Fukuda S., Furuno M., Hanagaki T., Hara A., Hashizume W.,
RA   Hayashida K., Hayatsu N., Hiramoto K., Hiraoka T., Hirozane T.,
RA   Hori F., Imotani K., Ishii Y., Itoh M., Kagawa I., Kasukawa T.,
RA   Katoh H., Kawai J., Kojima Y., Kondo S., Konno H., Kouda M., Koya S.,
RA   Kurihara C., Matsuyama T., Miyazaki A., Murata M., Nakamura M.,
RA   Nishi K., Nomura K., Numazaki R., Ohno M., Ohsato N., Okazaki Y.,
RA   Saito R., Saitoh H., Sakai C., Sakai K., Sakazume N., Sano H.,
RA   Sasaki D., Shibata K., Shinagawa A., Shiraki T., Sogabe Y., Tagami M.,
RA   Tagawa A., Takahashi F., Takaku-Akahira S., Takeda Y., Tanaka T.,
RA   Tomaru A., Toya T., Yasunishi A., Muramatsu M., Hayashizaki Y.;
RL   Submitted (APR-2002) to the EMBL/GenBank/DDBJ databases.
DR   EMBL; AK088796; BAC40578.1; -; mRNA.
DR   MGI; MGI:103199; Mcm4.
DR   GO; GO:0005634; C:nucleus; IEA.
DR   GO; GO:0005524; F:ATP binding; IEA.
DR   GO; GO:0003677; F:DNA binding; IEA.
DR   GO; GO:0008094; F:DNA-dependent ATPase activity; IEA.
DR   GO; GO:0006270; P:DNA replication initiation; IEA.
DR   InterPro; IPR001208; MCM.
DR   InterPro; IPR008047; MCM_4.
DR   Pfam; PF00493; MCM; 1.
DR   PRINTS; PR01657; MCMFAMILY.
DR   PRINTS; PR01660; MCMPROTEIN4.
DR   ProDom; PD001041; MCM; 1.
DR   SMART; SM00350; MCM; 1.
DR   PROSITE; PS00847; MCM_1; 1.
DR   PROSITE; PS50051; MCM_2; 1.
SQ   SEQUENCE   862 AA;  96736 MW;  516ACC1A3C6FB16E CRC64;

  Query Match             39.5%;  Score 1821;  DB 2;  Length 862;
  Best Local Similarity   43.7%;  Pred. No. 2.4e-92;
  Matches  401;  Conservative  154;  Mismatches  258;  Indels  104;  Gaps   18;

Qy         30 VTSPSFDNAASPVAGRRAVRQ---TPTSAVRRRGRETDSARRRRSRSRSLG--------- 77
              ::||    |::|    ||: |:    ||| ::|     :    ||||     | |
Db          1 MSSP----ASTP--SRRSSRRGRVTPTQSLRSEESRSSPNRRRRGEDSSTGELLPMPTSP 54

Qy         78 ----------NSVYSSP-----------YDAGTPGTPGTPVATPVYATPVGTPMGTP-SF 115
                        |:::|||             :|  :|  |          | ||        ||| |
Db         55 GADLQSPPAQNALFSSPPQMHSLAIPLDFDVSSPLT---------YGTPSSRVEGTPRSG 105

Qy        116 HRGTPQYKQRSELGSQGKPLHRRRRSQSREPGHRSPSREPSADGRPSESAEPDDTLGGEY 175
              ||||  :|| :|||  | |      :|                     || :|   |  :  |:
Db        106 VRGTP-VRQRPDLGSARKGLQVDLQS----------------DGAAAEDIVPSEQSLGQK 148

Qy        176 AYVWGTNVNIPDVLRAIRRFLHNY--------RSSAHDLNSK-YIQIIEETVEREEDTLN 226
                 :|||:||:        :|||    :          :    |:| : |            | ||
Db        149 LVIWGTDVNVATCKENFQRFLQCFTDPLAKEEENVGIDITQPLYMQQLGEINITGEPFLN 208

Qy        227 IDMSDIYDHDPDLYAKIVRYPLDIIPLLDTECQEVATSLLP--TFEKHIEARPFNLKASV 284
              ::    |        :||  :::  ||  ||  ::||    |       |:          |    |   |: ||||     :
Db        209 VNCEHIKSFSKNLYRQLISYPQEVIPTFDMAVNEIFFDRYPDSILEHQIQVRPFNALKTK 268

Qy        285 HMRELNPSDIDKLVSVKGMVIRCSSIIPEIKGAFFKCLVCGHSPPLVTVVKGRVEEPTRC 344
               || ||| |||:|::: |||||  | :|||: |||:| || |:      | : :||: || |
Db        269 SMRNLNPEDIDQLITISGMVIRTSQLIPEMQEAFFQCQVCAHT-TRVEIDRGRIAEPCSC 327

Qy        345 EKPECAARNAMSLIHNRCTFANKQIVRLQETPDAIPEGETPHTVSMCLYNTMVDAVKPGD 404
                   |     ::|:|||||   |::|:::|||:|: :| |:||||: :   :| :|| |:|||
Db        328 --VHCHTTHSMALIHNRSFFSDKQMIKLQESPEDMPAGQTPHTIVLFAHNDLVDKVQPGD 385

Qy        405 RIEVTGVFKAMAVRVGPNQRTLRALYKTYIDCVHVKKSDRGRLQTEDPMEMDKENDMYAG 464
              |: |||:::|: :|| |        ::::|||:|| :| :|:|   ||                         |
Db        386 RVNVTGIYRAVPIRVNPRVSNVKSVYKTHIDVIHYRKTDAKRLH--------------G 430

Qy        465 YHESDTSEAANEAKIQKLKELSKLPGIYDRLSRSLAPSIWELEDIKKGLLCQLFGGKAKK 524
                   |     :   :| :::  |||||:  | ||:||: |||||:| |||||:| |||||  |
Db        431 LDEEAEQKLFSEKRVKLLKELSRKPDIYERLASALAPSIYEHEDIKKGILLQLFGGTRKD 490

Qy        525 I--PSGASFRGDINVLLVGDPGTSKSQLLQYVHKIAPRGIYTSGRGSSAVGLTAYVTKDP 582
                    || :||:|| ||||||||||||||: : ||| ||||:|||||||||||| |||
Db        491 FSHTGRGKFRAEINILLCGDPGTSKSQLLQYVYNLVPRGQYTSGKGSSAVGLTAYVMKDP 550

Qy        583 ETRETVLESGALVLSDRGICCIDEFDKMSDNARSMLHEVMEQQTVSVAKGGIIASLNART 642
              ||||: ||::||||||| ||||||||||||::: ||:|||||||||:|:|| |||  |||||
Db        551 ETRQLVLQTGALVLSDNGICCIDEFDKMNESTRSVLHEVMEQQTLSIAKAGIICQLNART 610

Qy        643 SVLACANPSGSRYNARLSVIDNIQLPPTLLSRFDLIYLMLDKPDEQNDRRLARHLVALHY 702
              |||| |||    |::|  : :  |:||||| |||||||||:||||  ||  ||||| |||:|:|
Db        611 SVLAAANPIESQWNPKKTTIENIQLPHTLLSRFDLIFLMLDPQDEAYDRRLAHHLVSLYY 670

Qy        703 ENYEVSKQDALDLQTLTAYITYARQHVHPTLSDEAAEDLINGYVEMRQKGNFPGSSKKVI 762
              :: |  :::  ||:   |   || ||    :  | ||:||:: || :| ||:          |||:  ::
Db        671 QSEEQVEEEFLDMAVLKDYIAYAHSTIMPRLSEEASQALIEAYVNMRK----IGSSRGMV 726

Qy        763 TATPRQLESMIRISEALARMRFSEVVEKVDAAEAVRLLDVALQQSATDHATGTIDMDLIT 822
              :| ||||||:||:||:::|| |::||| |:  :| ||:||:| || ||:||||| || ||: ::|
Db        727 SAYPRQLESLIRLAEAHAKVRFSNKVEAIDVEEAKRLHREALKQSATDPRTGIVDISILT 786

Qy        823 TGVSASERIRRANLLAALRELIADKISPGSSSGLKTSQLLEDIRSQSSVDVSLQDIKNAL 882
              ||:||: | |:   |  |||:||   :| | :  ||  || ||||  ||      ::         :||
Db        787 TGMSATSRKRKEELAEALRKLI---LSKGKTPALKYQQLFEDIRGQSDTAITKDMFEEAL 843

Qy        883 GSLQGEGFLTVHGDIVK 899
               :|   : |||| |  |:
Db        844 RALADDDFLTVTGKTVR 860


RESULT 14
Q6GL41_XENTR
ID   Q6GL41_XENTR PRELIMINARY;      PRT;   863 AA.
AC   Q6GL41;
DT   05-JUL-2004 (TrEMBLrel. 27, Created)
DT   05-JUL-2004 (TrEMBLrel. 27, Last sequence update)
DT   05-JUL-2004 (TrEMBLrel. 27, Last annotation update)
DE   MCM4 minichromosome maintenance deficient 4.
GN   Name=mcm4-prov;
OS   Xenopus tropicalis (Western clawed frog) (Silurana tropicalis).
OC   Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC   Amphibia; Batrachia; Anura; Mesobatrachia; Pipoidea; Pipidae;
OC   Xenopodinae; Xenopus; Silurana.
OX   NCBI_TaxID=8364;
RN   [1]
RP   NUCLEOTIDE SEQUENCE.
RC   TISSUE=Embryo;
RX   MEDLINE=22388257; PubMed=12477932; DOI=10.1073/pnas.242603899;
RA   Strausberg R.L., Feingold E.A., Grouse L.H., Derge J.G.,
RA   Klausner R.D., Collins F.S., Wagner L., Shenmen C.M., Schuler G.D.,
RA   Altschul S.F., Zeeberg B., Buetow K.H., Schaefer C.F., Bhat N.K.,
RA   Hopkins R.F., Jordan H., Moore T., Max S.I., Wang J., Hsieh F.,
RA   Diatchenko L., Marusina K., Farmer A.A., Rubin G.M., Hong L.,
RA   Stapleton M., Soares M.B., Bonaldo M.F., Casavant T.L., Scheetz T.E.,
```

```
RA   Brownstein M.J., Usdin T.B., Toshiyuki S., Carninci P., Prange C.,
RA   Raha S.S., Loquellano N.A., Peters G.J., Abramson R.D., Mullahy S.J.,
RA   Bosak S.A., McEwan P.J., McKernan K.J., Malek J.A., Gunaratne P.H.,
RA   Richards S., Worley K.C., Hale S., Garcia A.M., Gay L.J., Hulyk S.W.,
RA   Villalon D.K., Muzny D.M., Sodergren E.J., Lu X., Gibbs R.A.,
RA   Fahey J., Helton E., Ketteman M., Madan A., Rodrigues S., Sanchez A.,
RA   Whiting M., Madan A., Young A.C., Shevchenko Y., Bouffard G.G.,
RA   Blakesley R.W., Touchman J.W., Green E.D., Dickson M.C.,
RA   Rodriguez A.C., Grimwood J., Schmutz J., Myers R.M.,
RA   Butterfield Y.S.N., Krzywinski M.I., Skalska U., Smailus D.E.,
RA   Schnerch A., Schein J.E., Jones S.J.M., Marra M.A.;
RT   "Generation and initial analysis of more than 15,000 full-length human
RT   and mouse cDNA sequences.";
RL   Proc. Natl. Acad. Sci. U.S.A. 99:16899-16903(2002).
RN   [2]
RP   NUCLEOTIDE SEQUENCE.
RC   TISSUE=Embryo;
RA   Klein S., Gerhard D.S.;
RL   Submitted (JUN-2004) to the EMBL/GenBank/DDBJ databases.
DR   EMBL; BC074670; AAH74670.1; -; mRNA.
DR   Ensembl; ENSXETG00000004854; Xenopus tropicalis.
DR   GO; GO:0005634; C:nucleus; IEA.
DR   GO; GO:0005524; F:ATP binding; IEA.
DR   GO; GO:0003677; F:DNA binding; IEA.
DR   GO; GO:0008094; F:DNA-dependent ATPase activity; IEA.
DR   GO; GO:0006270; P:DNA replication initiation; IEA.
DR   InterPro; IPR001208; MCM.
DR   InterPro; IPR008047; MCM_4.
DR   Pfam; PF00493; MCM; 1.
DR   PRINTS; PR01657; MCMFAMILY.
DR   PRINTS; PR01660; MCMPROTEIN4.
DR   ProDom; PD001041; MCM; 1.
DR   SMART; SM00350; MCM; 1.
DR   PROSITE; PS00847; MCM_1; 1.
DR   PROSITE; PS50051; MCM_2; 1.
SQ   SEQUENCE   863 AA;  97196 MW;  2AAFF32FE85BE97B CRC64;

  Query Match             39.4%;  Score 1817.5;  DB 2;  Length 863;
  Best Local Similarity   44.2%;  Pred. No. 3.7e-92;
  Matches  402;  Conservative  145;  Mismatches  247;  Indels  116;  Gaps   19;

Qy         51 TPTSAVRRRGRETD-------------SARRRRSRSRSLG-------------------- 77
              ||:    :|||  ::                    | |||    | |:|
Db          7 TPSRRRNKRGRGSNPPTPHGEEVQSPPSQRRRTEDSTSIGELLPMPTSPSGDLQSPSGQE 66

Qy         78 ----------NSVYSSPYDAGTPGTPGTPVATPVYATPVGTPMGTP-SFHRGTPQYKQRS 126
                        :|    |   |   :|  |            | ||        ||| |  ||||   :||
Db         67 LMFSSPAPSRHSALQSELDLSSPLT---------YGTPSSRVEGTPRSGIRGTPA-RQRP 116

Qy        127 ELGSQGK----PLHRRRRSQSREPGHRSPSREPSADGRPSESAEPDDTLGGEYAYVWGTN 182
              :|||   |        ||                        | :|:|:      |      :  :|   |:     :|||:
Db        117 DLGSARKVKQVDLH---------------SDQPAAE----ELVTSEQSL-GQKLVIWGTD 156

Qy        183 VNIPDVLRAIRRFLHNY--------RSSAHDLNSK-YIQIIEETVEREEDTLNIDMSDIY 233
              ||:          :|:|:                :  |||     |:| :||        ||:|     :
Db        157 VNVATCKEKFQRFVQRFIDPLAKEEENVGLDLNEPIYMQRLEEINVVGEPFLNVDCDHLR 216

Qy        234 DHDPDLYAKIVRYPLDIIPLLDTECQEVATSLLP--TFEKHIEARPFNLKASVHMRELNP 291
               | |||  ::| ||  ::||   |        |:       |       |    |: ||:|      : :|| |||
Db        217 SFDQDLYRQLVCYPQEVIPTFDMAANEIFFERYPDSILEHQIQVRPYNALKTRNMRSLNP 276

Qy        292 SDIDKLVSVKGMVIRCSSIIPEIKGAFFKCLVCGHSPPLVTVVKGRVEEPTRCEKPECAA 351
               |||:|:::  |||||  |  ||||::  |||||  ||     :      |  :  :||:  ||:  |:     |
Db        277 EDIDQLITISGMVIRTSQIIPEMQEAFFKCQVCAFT-TRVEIDRGRISEPSVCK--HCNT 333

Qy        352 RNAMSLIHNRCTFANKQIVRLQETPDAIPEGETPHTVSMCLYNTMVDAVKPGDRIEVTGV 411
              ::|:|||||    |::||:::|||:|:   :|   |:||||     :     :|  :||  |:||||:  ||:  |||:
Db        334 THSMALIHNRSMFSDKQMIKLQESPEDMPAGQTPHTTILYGHNDLVDKVQPGDRVNVTGI 393

Qy        412 FKAMAVRVGPNQRTLRALYKTYIDCVHVKKSDRGRLQTEDPMEMDKENDMYAGYHESDTS 471
              ::|:  :|| |     | ::::||||:||   :|  :|:|        ||          |      |  |     |:
Db        394 YRAVPIRVNPRVRNVKSVYKTHIDVIHYRKTDAKRLHGIDE---DTEQKMF--------- 441

Qy        472 EAANEAKIQKLKELSKLPGIYDRLSRSLAPSIWELEDIKKGLLCQLFGGKAKKI--PSGA 529
                 | ::   ||||:   |  ||:||:  :|||||:| ||||||:| |||||    |
Db        442 ---TEERVAMLKELAAKPDIYERLASALAPSIYEHEDIKKGILLQLFGGTRKDFSHTGRG 498

Qy        530 SFRGDINVLLVGDPGTSKSQLLQYVHKIAPRGIYTSGRGSSAVGLTAYVTKDPETRETVL 589
                || ::|:|| |||||||||||||:  :  ||| ||||:||||||||||| ||||||: ||
Db        499 KFRAEVNILLCGDPGTSKSQLLQYVYNLVPRGQYTSGKGSSAVGLTAYVMKDPETRQLVL 558

Qy        590 ESGALVLSDRGICCIDEFDKMSDNARSMLHEVMEQQTVSVAKGGIIASLNARTSVLACAN 649
              ::||||||| ||||||||||||:::  |||:|||||||||||:|:|| |||   ||||||||| ||
Db        559 QTGALVLSDNGICCIDEFDKMNESTRSVLHEVMEQQTLSIAKAGIICQLNARTSVLAAAN 618

Qy        650 PSGSRYNARLSVIDNIQLPPTLLSRFDLIYLMLDKPDEQNDRRLARHLVALHYENYEVSK 709
               |    |::| :  : |:||||| |||||||||:||||  ||   |||||  |||||:|:: |   |
Db        619 PVESQWNPKKTTIENIQLPHTLLSRFDLIFLMLDPQDETYDRRLAHHLVALYYQSEEQMK 678

Qy        710 QDALDLQTLTAYITYARQHVHPTLSDEAAEDLINGYVEMRQKGNFPGSSKKVITATPRQL 769
              :: ||:    |    || |||  :|:|  |  :||::  ||   ||:||:     || :  :::| ||||
Db        679 EEHLDMAVLKDYIAYARTYVNPRLGEEASQALIEAYVDMRK----IGSGRGMVSAYPRQL 734

Qy        770 ESMIRISEALARMRFSEVVEKVDAAEAVRLLDVALQQSATDHATGTIDMDLITTGVSASE 829
              ||:||:|||  |::|||  ||  :|    || ||    ||:|||||  || :|:  ::|||:||:
Db        735 ESLIRLSEAHAKVRFSSKVETIDVEEAKRLHREALKQSATDPRTGIVDISILTTGMSATA 794

Qy        830 RIRRANLLAALRELIADKISPGSSSGLKTSQLLEDIRSQSSVDVSLQDIKNALGSLQGEG 889
               | |:   |      |:||    |   :    ||  || |:|   ||    |:          || :|    |
Db        795 RKRKEELAQVLKKLIQSK---GKTPALKYQQLFEDLRGQSDAAVTKDMFDEALHALADED 851

Qy        890 FLTVHGDIVK 899
               :|||  |   |:
Db        852 YLTVTGKTVR 861

RESULT 15
Q8C1Z0_MOUSE
ID   Q8C1Z0_MOUSE PRELIMINARY;      PRT;   862 AA.
AC   Q8C1Z0;
DT   01-MAR-2003 (TrEMBLrel. 23, Created)
DT   01-MAR-2003 (TrEMBLrel. 23, Last sequence update)
DT   01-MAR-2004 (TrEMBLrel. 26, Last annotation update)
DE   Mus musculus brain CRL-1443 BC3H1 cDNA, RIKEN full-length enriched
DE   library, clone:G430050N09 product:mini chromosome maintenance
DE   deficient 4 homolog (S. cerevisiae), full insert sequence.
GN   Name=Mcm4;
OS   Mus musculus (Mouse).
OC   Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC   Mammalia; Eutheria; Euarchontoglires; Glires; Rodentia; Sciurognathi;
OC   Muridae; Murinae; Mus.
OX   NCBI_TaxID=10090;
RN   [1]
RP   NUCLEOTIDE SEQUENCE.
RC   STRAIN=C3H; TISSUE=Brain;
RX   MEDLINE=99279253; PubMed=10349636; DOI=10.1016/S0076-6879(99)03004-9;
RA   Carninci P., Hayashizaki Y.;
RT   "High-efficiency full-length cDNA cloning.";
RL   Meth. Enzymol. 303:19-44(1999).
RN   [2]
RP   NUCLEOTIDE SEQUENCE.
RC   STRAIN=C3H; TISSUE=Brain;
RX   MEDLINE=21085660; PubMed=11217851; DOI=10.1038/35055500;
RA   Kawai J., Shinagawa A., Shibata K., Yoshino M., Itoh M., Ishii Y.,
RA   Arakawa T., Hara A., Fukunishi Y., Konno H., Adachi J., Fukuda S.,
RA   Aizawa K., Izawa M., Nishi K., Kiyosawa H., Kondo S., Yamanaka I.,
RA   Saito T., Okazaki Y., Gojobori T., Bono H., Kasukawa T., Saito R.,
RA   Kadota K., Matsuda H.A., Ashburner M., Batalov S., Casavant T.,
RA   Fleischmann W., Gaasterland T., Gissi C., King B., Kochiwa H.,
RA   Kuehl P., Lewis S., Matsuo Y., Nikaido I., Pesole G., Quackenbush J.,
RA   Schriml L.M., Staubli F., Suzuki R., Tomita M., Wagner L., Washio T.,
RA   Sakai K., Okido T., Furuno M., Aono H., Baldarelli R., Barsh G.,
RA   Blake J., Boffelli D., Bojunga N., Carninci P., de Bonaldo M.F.,
RA   Brownstein M.J., Bult C., Fletcher C., Fujita M., Gariboldi M.,
RA   Gustincich S., Hill D., Hofmann M., Hume D.A., Kamiya M., Lee N.H.,
RA   Lyons P., Marchionni L., Mashima J., Mazzarelli J., Mombaerts P.,
```

```
RA   Nordone P., Ring B., Ringwald M., Rodriguez I., Sakamoto N.,
RA   Sasaki H., Sato K., Schoenbach C., Seya T., Shibata Y., Storch K.-F.,
RA   Suzuki H., Toyo-oka K., Wang K.H., Weitz C., Whittaker C., Wilming L.,
RA   Wynshaw-Boris A., Yoshida K., Hasegawa Y., Kawaji H., Kohtsuki S.,
RA   Hayashizaki Y.;
RT   "Functional annotation of a full-length mouse cDNA collection.";
RL   Nature 409:685-690(2001).
RN   [3]
RP   NUCLEOTIDE SEQUENCE.
RC   STRAIN=C3H; TISSUE=Brain;
RA   The FANTOM Consortium,
RA   the RIKEN Genome Exploration Research Group Phase I & II Team;
RT   "Analysis of the mouse transcriptome based on functional annotation of
RT   60,770 full-length cDNAs.";
RL   Nature 420:563-573(2002).
RN   [4]
RP   NUCLEOTIDE SEQUENCE.
RC   STRAIN=C3H; TISSUE=Brain;
RX   MEDLINE=20499374; PubMed=11042159; DOI=10.1101/gr.145100;
RA   Carninci P., Shibata Y., Hayatsu N., Sugahara Y., Shibata K., Itoh M.,
RA   Konno H., Okazaki Y., Muramatsu M., Hayashizaki Y.;
RT   "Normalization and subtraction of cap-trapper-selected cDNAs to
RT   prepare full-length cDNA libraries for rapid discovery of new genes.";
RL   Genome Res. 10:1617-1630(2000).
RN   [5]
RP   NUCLEOTIDE SEQUENCE.
RC   STRAIN=C3H; TISSUE=Brain;
RX   MEDLINE=20530913; PubMed=11076861; DOI=10.1101/gr.152600;
RA   Shibata K., Itoh M., Aizawa K., Nagaoka S., Sasaki N., Carninci P.,
RA   Konno H., Akiyama J., Nishi K., Kitsunai T., Tashiro H., Itoh M.,
RA   Sumi N., Ishii Y., Nakamura S., Hazama M., Nishine T., Harada A.,
RA   Yamamoto R., Matsumoto H., Sakaguchi S., Ikegami T., Kashiwagi K.,
RA   Fujiwake S., Inoue K., Togawa Y., Izawa M., Ohara E., Watahiki M.,
RA   Yoneda Y., Ishikawa T., Ozawa K., Tanaka T., Matsuura S., Kawai J.,
RA   Okazaki Y., Muramatsu M., Inoue Y., Kira A., Hayashizaki Y.;
RT   "RIKEN integrated sequence analysis (RISA) system-384-format
RT   sequencing pipeline with 384 multicapillary sequencer.";
RL   Genome Res. 10:1757-1771(2000).
RN   [6]
RP   NUCLEOTIDE SEQUENCE.
RC   STRAIN=C3H; TISSUE=Brain;
RA   Adachi J., Aizawa K., Akimura T., Arakawa T., Bono H., Carninci P.,
RA   Fukuda S., Furuno M., Hanagaki T., Hara A., Hashizume W.,
RA   Hayashida K., Hayatsu N., Hiramoto K., Hiraoka T., Hirozane T.,
RA   Hori F., Imotani K., Ishii Y., Itoh M., Kagawa I., Kasukawa T.,
RA   Katoh H., Kawai J., Kojima Y., Kondo S., Konno H., Kouda M., Koya S.,
RA   Kurihara C., Matsuyama T., Miyazaki A., Murata M., Nakamura M.,
RA   Nishi K., Nomura K., Numazaki R., Ohno M., Ohsato N., Okazaki Y.,
RA   Saito R., Saitoh H., Sakai C., Sakai K., Sakazume N., Sano H.,
RA   Sasaki D., Shibata K., Shinagawa A., Shiraki T., Sogabe Y., Tagami M.,
RA   Tagawa A., Takahashi F., Takaku-Akahira S., Takeda Y., Tanaka T.,
RA   Tomaru A., Toya T., Yasunishi A., Muramatsu M., Hayashizaki Y.;
RL   Submitted (APR-2002) to the EMBL/GenBank/DDBJ databases.
DR   EMBL; AK089999; BAC41036.1; -; mRNA.
DR   MGI; MGI:103199; Mcm4.
DR   GO; GO:0005634; C:nucleus; IEA.
DR   GO; GO:0005524; F:ATP binding; IEA.
DR   GO; GO:0003677; F:DNA binding; IEA.
DR   GO; GO:0008094; F:DNA-dependent ATPase activity; IEA.
DR   GO; GO:0006270; P:DNA replication initiation; IEA.
DR   InterPro; IPR001208; MCM.
DR   InterPro; IPR008047; MCM_4.
DR   Pfam; PF00493; MCM; 1.
DR   PRINTS; PR01657; MCMFAMILY.
DR   PRINTS; PR01660; MCMPROTEIN4.
DR   ProDom; PD001041; MCM; 1.
DR   SMART; SM00350; MCM; 1.
DR   PROSITE; PS50051; MCM_2; 1.
SQ   SEQUENCE   862 AA;  96752 MW;  FF2FA597021EB8E6 CRC64;

  Query Match             39.4%;  Score 1817;  DB 2;  Length 862;
  Best Local Similarity   43.6%;  Pred. No. 4e-92;
  Matches  400;  Conservative  155;  Mismatches  258;  Indels  104;  Gaps   18;

Qy         30 VTSPSFDNAASPVAGRRAVRQ---TPTSAVRRRGRETDSARRRRSRSRSLG--------- 77
              ::||    |::|    ||: |:    ||| ::|        :    ||||     | |
Db          1 MSSP----ASTP--SRRSSRRGRVTPTQSLRSEESRSSPNRRRRGEDSSTGELLPMPTSP 54

Qy         78 ----------NSVYSSP-----------YDAGTPGTPGTPVATPVYATPVGTPMGTP-SF 115
                        |:::|||              :|  :| |          | ||     ||| |
Db         55 GADLQSPPAQNALFSSPPQMHSLAIPLDFDVSSPLT---------YGTPSSRVEGTPRSG 105

Qy        116 HRGTPQYKQRSELGSQGKPLHRRRRSQSREPGHRSPSREPSADGRPSESAEPDDTLGGEY 175
              ||||  :|| :|||  | |      :|                    || :|   |  :  |:
Db        106 VRGTP-VRQRPDLGSARKGLQVDLQS----------------DGAAAEDIVPSEQSLGQK 148

Qy        176 AYVWGTNVNIPDVLRAIRRFLHNY--------RSSAHDLNSK-YIQIIEETVEREEDTLN 226
                :|||:||:         :|||    :              :    |:    |:| :      |  ||
Db        149 LVIWGTDVNVATCKENFQRFLQCFTDPLAQEEENVGIDITQPLYMQQLGEINITGEPFLN 208

Qy        227 IDMSDIYDHDPDLYAKIVRYPLDIIPLLDTECQEVATSLLP--TFEKHIEARPFNLKASV 284
              ::    |       :||  :::  || ::||   |       |:       |      |  |:  ||||     :
Db        209 VNCEHIKSFSKNLYRQLISYPQEVIPTFDMAVNEIFFDRYPDSILEHQIQVRPFNALKTK 268

Qy        285 HMRELNPSDIDKLVSVKGMVIRCSSIIPEIKGAFFKCLVCGHSPPLVTVVKGRVEEPTRC 344
               || ||| |||:|:::  |||||  | :|||::  |||:|  || |:    | : :||: || |
Db        269 SMRNLNPEDIDQLITISGMVIRTSQLIPEMQEAFFQCQVCAHT-TRVEIDRGRIAEPCSC 327

Qy        345 EKPECAARNAMSLIHNRCTFANKQIVRLQETPDAIPEGETPHTVSMCLYNTMVDAVKPGD 404
                   |     ::|:|||||   |::||:::|||:|: :|  |:||||: :   :| :|| |:|||
Db        328 --VHCHTTHSMALIHNRSFFSDKQMIKLQESPEDMPAGQTPHTIVLFAHNDLVDKVQPGD 385

Qy        405 RIEVTGVFKAMAVRVGPNQRTLRALYKTYIDCVHVKKSDRGRLQTEDPMEMDKENDMYAG 464
              |: |||:::|: :|| |     ::::|||:|| :| :|:| ||                   |
Db        386 RVNVTGIYRAVPIRVNPRVSNVKSVYKTHIDVIHYRKTDAKRLH--------------G 430

Qy        465 YHESDTSEAANEAKIQKLKELSKLPGIYDRLSRSLAPSIWELEDIKKGLLCQLFGGKAKK 524
                |       :   :| :::  ||||: | ||:||: :|||||:| ||||||:| |||||    |
Db        431 LDEEAEQKLFSEKRVKLLKELSRKPDIYERLASALAPSIYEHEDIKKGILLQLFGGTRKD 490

Qy        525 I--PSGASFRGDINVLLVGDPGTSKSQLLQYVHKIAPRGIYTSGRGSSAVGLTAYVTKDP 582
                           || :||:|| ||||||||||||||: :  |||  ||||:||||||||||| |||
Db        491 FSHTGRGKFRAEINILLCGDPGTSKSQLLQYVYNLVPRGQYTSGKGSSAVGLTAYVMKDP 550

Qy        583 ETRETVLESGALVLSDRGICCIDEFDKMSDNARSMLHEVMEQQTVSVAKGGIIASLNART 642
              ||||: ||::||||||| ||||||:||||::: ||:||||||||||:|:|| |||  |||||
Db        551 ETRQLVLQTGALVLSDNGICCIDKFDKMNESTRSVLHEVMEQQTLSIAKAGIICQLNART 610

Qy        643 SVLACANPSGSRYNARLSVIDNIQLPPTLLSRFDLIYLMLDKPDEQNDRRLARHLVALHY 702
              |||| |||   |::|  : :  |:||||  ||||||||:||||  ||  |||||  |||:|:|
Db        611 SVLAAANPIESQWNPKKTTIENIQLPHTLLSRFDLIFLMLDPQDEAYDRRLAHHLVSLYY 670

Qy        703 ENYEVSKQDALDLQTLTAYITYARQHVHPTLSDEAAEDLINGYVEMRQKGNFPGSSKKVI 762
              :: |   ::: ||:   |  ||| ||      :  | ||:||::  ||   || ||:     |||: ::
Db        671 QSEEQVEEEFLDMAVLKDYIAYAHSTIMPRLSEEASQALIEAYVNMRK----IGSSRGMV 726

Qy        763 TATPRQLESMIRISEALARMRFSEVVEKVDAAEAVRLLDVALQQSATDHATGTIDMDLIT 822
              :| ||||||:||::||:| |::|||   || :|   || ||     ||:|||||  || :|: ::|
Db        727 SAYPRQLESLIRLAEAHAKVRFSNKVEAIDVEEAKRLHREALKQSATDPRTGMVDISILT 786

Qy        823 TGVSASERIRRANLLAALRELIADKISPGSSSGLKTSQLLEDIRSQSSVDVSLQDIKNAL 882
              ||:||: | |:  |  |||:||   :| |  :  ||  || |||| ||    ::    :  ||
Db        787 TGMSATSRKRKEELAEALRKLI---LSKGKTPALKYQQLFEDIRGQSDTAITKDMFEEAL 843

Qy        883 GSLQGEGFLTVHGDIVK 899
               :|  : |||| |  |:
Db        844 RALADDDFLTVTGKTVR 860

Search completed: December  7, 2005, 17:42:56
Job time : 247 secs
```

```
                       GenCore version 5.1.6
              Copyright (c) 1993 - 2005  Compugen Ltd.

OM protein - nucleic search, using frame_plus_p2n model

Run on:          December  6, 2005, 09:08:24 ; Search time 7664 Seconds
                                            (without alignments)
                                            6682.666 Million cell updates/sec

Title:           US-10-768-511-8
Perfect score:   4609
Sequence:        1 MENNDALDIGAVSSPYPSQS..........LGSLQGEGFLTVHGDIVKRV 901

Scoring table:   BLOSUM62
                 Xgapop 10.0 , Xgapext  0.5
                 Ygapop 10.0 , Ygapext  0.5
                 Fgapop  6.0 , Fgapext  7.0
                 Delop   6.0 , Delext   7.0

Searched:        5883141 seqs, 28421725653 residues

Total number of hits satisfying chosen parameters:      11766282

Minimum DB seq length: 0
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%
                 Maximum Match 100%
                 Listing first 45 summaries

Command line parameters:
-MODEL=frame+_p2n.model -DEV=xlh
-Q=/cgn2_1/USPTO_spool/US10768511/runat_02122005_102334_18404/app_query.fasta_1.1095
-DB=GenEmbl -QFMT=fastap -SUFFIX=rge -MINMATCH=0.1 -LOOPCL=0 -LOOPEXT=0
-UNITS=bits -START=1 -END=-1 -MATRIX=blosum62 -TRANS=human40.cdi -LIST=45
-DOCALIGN=200 -THR_SCORE=pct -THR_MAX=100 -THR_MIN=0 -ALIGN=15 -MODE=LOCAL
-OUTFMT=pto -NORM=ext -HEAPSIZE=500 -MINLEN=0 -MAXLEN=2000000000
-USER=US10768511_@CGN_1_1_5186_@runat_02122005_102334_18404 -NCPU=6 -ICPU=3
-NO_MMAP -LARGEQUERY -NEG_SCORES=0 -WAIT -DSPBLOCK=100 -LONGLOG
-DEV_TIMEOUT=120 -WARN_TIMEOUT=30 -THREADS=1 -XGAPOP=10 -XGAPEXT=0.5 -FGAPOP=6
-FGAPEXT=7 -YGAPOP=10 -YGAPEXT=0.5 -DELOP=6 -DELEXT=7

Database :      GenEmbl:*
                1:  gb_ba:*
                2:  gb_in:*
                3:  gb_env:*
                4:  gb_om:*
                5:  gb_ov:*
                6:  gb_pat:*
                7:  gb_ph:*
                8:  gb_pr:*
                9:  gb_ro:*
                10:  gb_sts:*
                11:  gb_sy:*
                12:  gb_un:*
                13:  gb_vi:*
                14:  gb_htg:*
                15:  gb_pl:*

     Pred. No. is the number of results predicted by chance to have a
     score greater than or equal to the score of the result being printed,
     and is derived by analysis of the total score distribution.

                               SUMMARIES
                   %
Result            Query
  No.    Score   Match Length DB  ID                         Description
-----------------------------------------------------------------------------
    1     4609   100.0    4348  6  AR490068                   AR490068 Sequence
    2     4609   100.0    4348  6  AX281543                   AX281543 Sequence
c   3     2016    43.7 110000  14  CR954214_1                  Continuation (2 of
    4   1937.5    42.0 110000  15  AP008207_201                Continuation (202
    5   1937.5    42.0 187471  15  AP004232                   AP004232 Oryza sat
c   6     1909    41.4 153751   2  AC116551                  AC116551 Dictyoste
    7     1883    40.9 114418  15  AC007047                   AC007047 Arabidops
    8     1882    40.8   3667  15  SPCDC21                    X58824 Schizosacch
    9   1858.5    40.3 116616  15  AC149637                   AC149637 Medicago
   10     1833    39.8   5392  15  SCU14731                   U14731 Saccharomyc
   11     1833    39.8  27166  15  SC9531X                    Z49919 S.cerevisia
   12     1833    39.8 165536  15  SCCHRXVI                   Z71255 S.cerevisia
   13   1822.5    39.5   3221   9  MUSMCDC21                 D26089 Mus musculu
   14   1821.5    39.5   3283   9  AK220442                  AK220442 Mus muscu
c  15   1819.5    39.5 110000  15  CR380957_02                 Continuation (3 of
   16   1817.5    39.4   3273   6  CQ981174                  CQ981174 Sequence
   17   1817.5    39.4   3273   6  AX333086                  AX333086 Sequence
   18   1817.5    39.4   3273   6  AX333087                  AX333087 Sequence
   19   1817.5    39.4   3273   6  AX333494                  AX333494 Sequence
   20   1817.5    39.4   3273   8  HSP1CDC21                 X74794 H.sapiens P
   21   1817.5    39.4   3273  10  G28594                     G28594 human STS S
   22   1817.5    39.4   3290   5  BC074670                  BC074670 Xenopus t
   23     1815    39.4   2851   8  BC031061                  BC031061 Homo sapi
   24   1812.5    39.3   3360   9  BC013094                  BC013094 Mus muscu
   25   1811.5    39.3   2860   6  BD159700                  BD159700 Primer fo
   26   1811.5    39.3   2860   6  AX882395                  AX882395 Sequence
   27   1811.5    39.3   2860   8  AK022899                  AK022899 Homo sapi
   28     1810    39.3   2970   5  XLU46131                  U46131 Xenopus lae
   29     1810    39.3   3048   5  BC072870                  BC072870 Xenopus l
   30   1809.5    39.3 110000  15  CR382137_13                 Continuation (14 o
   31   1808.5    39.2   2893   5  BC083031                  BC083031 Xenopus l
   32     1802    39.1   3051   5  XLU29178                  U29178 Xenopus lae
   33     1802    39.1 105066  14  AP006498_8                  Continuation (9 of
c  34   1793.5    38.9 110000  15  CR382123_15                 Continuation (16 o
   35     1792    38.9 110000  15  AE016817_06                 Continuation (7 of
   36   1791.5    38.9   3042   5  BC065958                  BC065958 Danio rer
   37     1791    38.9   3394   6  AX281814                  AX281814 Sequence
   38   1787.5    38.8   3087   5  XLU44049                  U44049 Xenopus lae
   39     1787    38.8 110000  15  CR382132_07                 Continuation (8 of
   40   1779.5    38.6   2556   6  CQ717883                  CQ717883 Sequence
   41   1768.5    38.4   3040   5  BC056514                  BC056514 Danio rer
   42     1752    38.0   2733   6  AX489180                  AX489180 Sequence
   43   1746.5    37.9 244538  14  AC129365                   AC129365 Rattus no
   44   1711.5    37.1 241140  14  AC119094                   AC119094 Rattus no
   45   1706.5    37.0   2776   2  AK114961                  AK114961 Ciona int

                              ALIGNMENTS

RESULT 1
AR490068
LOCUS       AR490068                4348 bp    DNA     linear   PAT 15-MAY-2004
DEFINITION  Sequence 5 from patent US 6710229.
ACCESSION   AR490068
VERSION     AR490068.1  GI:47257191
KEYWORDS    .
SOURCE      Unknown.
  ORGANISM  Unknown.
            Unclassified.
REFERENCE   1  (bases 1 to 4348)
  AUTHORS   da Costa e Silva,O., Bohnert,H.J., van Thielen,N., Chen,R. and
            Sarria-Millan,R.
  TITLE     Cell cycle stress-related proteins and methods of use in plants
  JOURNAL   Patent: US 6710229-A 5 23-MAR-2004;
            BASF Plant Science GmbH; Ludwigshafen;
            WOX;
FEATURES             Location/Qualifiers
     source          1..4348
                     /organism="unknown"
                     /mol_type="genomic DNA"
ORIGIN

Alignment Scores:
Pred. No.:             1.23e-162       Length:         4348
Score:                 4609.00         Matches:        901
Percent Similarity:    100.00%         Conservative:   0
```

```
Best Local Similarity:  100.00%     Mismatches:   0
Query Match:            100.00%     Indels:       0
DB:                     6           Gaps:         0

US-10-768-511-8 (1-901) x AR490068 (1-4348)

Qy       1 MetGluAsnAsnAspAlaLeuAspIleGlyAlaValSerSerProTyrProSerGlnSer 20
Db    1554 ATGGAAAATAATGATGCACTTGACATTGGAGCCGTGTCGTCCCCATATCCTTCGCAATCT 1613
Qy      21 GluGlyValSerThrProLeuProGlnValThrSerProSerPheAspAsnAlaAlaSer 40
Db    1614 GAAGGAGTGTCTACGCCATTGCCGCAAGTAACATCACCGAGCTTCGACAATGCAGCCTCA 1673
Qy      41 ProValAlaGlyArgArgAlaValArgGlnThrProThrSerAlaValArgArgArgGly 60
Db    1674 CCCGTGGCCGGGCGGAGGGCCGTACGGCAGACCCCTACATCTGCAGTTCGAAGGAGAGGG 1733
Qy      61 ArgGluThrAspSerAlaArgArgArgArgSerArgSerArgSerLeuGlyAsnSerVal 80
Db    1734 AGAGAAACGGATTCCGCTCGTCGTAGGAGGAGTCGATCTCGCAGTTTAGGCAATTCTGTT 1793
Qy      81 TyrSerSerProTyrAspAlaGlyThrProGlyThrProGlyThrProValAlaThrPro 100
Db    1794 TATAGTTCCCCTTACGATGCGGGGACTCCTGGAACTCCTGGAACTCCAGTGGCTACTCCG 1853
Qy     101 ValTyrAlaThrProValGlyThrProMetGlyThrProSerPheHisArgGlyThrPro 120
Db    1854 GTTTACGCTACCCCAGTCGGTACACCTATGGGTACCCCATCGTTCCATCGTGGCACGCCA 1913
Qy     121 GlnTyrLysGlnArgSerGluLeuGlySerGlnGlyLysProLeuHisArgArgArgArg 140
Db    1914 CAGTACAAACAGCGCAGTGAGCTTGGTTCCCAGGGGAAGCCTCTACATCGGAGACGTCGA 1973
Qy     141 SerGlnSerArgGluProGlyHisArgSerProSerArgGluProSerAlaAspGlyArg 160
Db    1974 TCTCAATCCAGAGAACCCGGGCATCGATCTCCTTCAAGGGAACCTAGTGCTGATGGGCGT 2033
Qy     161 ProSerGluSerAlaGluProAspAspThrLeuGlyGlyGluTyrAlaTyrValTrpGly 180
Db    2034 CCCTCTGAATCTGCTGAGCCAGATGACACTTTGGGTGGAGAATATGCTTATGTTTGGGGG 2093
Qy     181 ThrAsnValAsnIleProAspValLeuArgAlaIleArgArgPheLeuHisAsnTyrArg 200
Db    2094 ACGAATGTTAACATTCCAGATGTGCTTAGGGCGATTCGTCGATTTCTCCACAATTATCGT 2153
Qy     201 SerSerAlaHisAspLeuAsnSerLysTyrIleGlnIleIleGluGluThrValGluArg 220
Db    2154 TCGAGTGCTCATGATCTTAATTCCAAGTACATCCAGATCATAGAGGAGACTGTGGAGCGT 2213
Qy     221 GluGluAspThrLeuAsnIleAspMetSerAspIleTyrAspHisAspProAspLeuTyr 240
Db    2214 GAGGAGGATACTCTAAATATCGACATGTCAGACATTTATGACCATGATCCTGATCTATAC 2273
Qy     241 AlaLysIleValArgTyrProLeuAspIleIleProLeuLeuAspThrGluCysGlnGlu 260
Db    2274 GCAAAAATTGTTCGATACCCACTCGACATCATCCCCCTGTTGGACACTGAGTGTCAGGAA 2333
Qy     261 ValAlaThrSerLeuLeuProThrPheGluLysHisIleGluAlaArgProPheAsnLeu 280
Db    2334 GTTGCTACCTCTTTACTACCAACGTTTGAGAAGCATATTGAGGCCAGACCTTTCAATCTC 2393
Qy     281 LysAlaSerValHisMetArgGluLeuAsnProSerAspIleAspLysLeuValSerVal 300
Db    2394 AAAGCATCGGTGCACATGCGTGAACTCAACCCTTCAGATATAGACAAATTGGTTTCTGTT 2453
Qy     301 LysGlyMetValIleArgCysSerSerIleIleProGluIleLysGlyAlaPhePheLys 320
Db    2454 AAAGGAATGGTTATCCGGTGCAGTTCTATCATACCTGAAATTAAGGGGGCCTTCTTCAAA 2513
Qy     321 CysLeuValCysGlyHisSerProProLeuValThrValValLysGlyArgValGluGlu 340
Db    2514 TGTTTAGTGTGTGGTCACTCGCCTCCGCTAGTTACAGTTGTTAAAGGGCGGGTTGAGGAG 2573
Qy     341 ProThrArgCysGluLysProGluCysAlaAlaArgAsnAlaMetSerLeuIleHisAsn 360
Db    2574 CCAACAAGGTGTGAAAAGCCAGAATGTGCAGCACGGAATGCTATGTCTCTTATTCACAAT 2633
Qy     361 ArgCysThrPheAlaAsnLysGlnIleValArgLeuGlnGluThrProAspAlaIlePro 380
Db    2634 CGATGCACTTTTGCAAATAAGCAGATAGTGCGTCTTCAAGAAACTCCAGATGCCATTCCT 2693
Qy     381 GluGlyGluThrProHisThrValSerMetCysLeuTyrAsnThrMetValAspAlaVal 400
Db    2694 GAAGGAGAGACTCCACACACAGTCAGCATGTGTTTATACAACACTATGGTTGATGCTGTG 2753
Qy     401 LysProGlyAspArgIleGluValThrGlyValPheLysAlaMetAlaValArgValGly 420
Db    2754 AAGCCTGGAGATCGTATTGAGGTAACAGGAGTTTTCAAGGCCATGGCAGTTCGAGTTGGT 2813
Qy     421 ProAsnGlnArgThrLeuArgAlaLeuTyrLysThrTyrIleAspCysValHisValLys 440
Db    2814 CCGAATCAACGAACATTACGAGCATTGTATAAGACCTACATCGATTGCGTGCACGTCAAG 2873
Qy     441 LysSerAspArgGlyArgLeuGlnThrGluAspProMetGluMetAspLysGluAsnAsp 460
Db    2874 AAGTCTGACAGGGGTCGACTGCAAACTGAAGATCCTATGGAGATGGATAAGGAGAATGAT 2933
Qy     461 MetTyrAlaGlyTyrHisGluSerAspThrSerGluAlaAlaAsnGluAlaLysIleGln 480
Db    2934 ATGTATGCTGGGTATCATGAAAGTGATACTTCAGAAGCTGCTAATGAAGCAAAGATTCAA 2993
Qy     481 LysLeuLysGluLeuSerLysLeuProGlyIleTyrAspArgLeuSerArgSerLeuAla 500
Db    2994 AAACTTAAAGAGCTGTCCAAGCTCCCGGGCATTTATGATAGACTTTCAAGGTCGCTGGCT 3053
Qy     501 ProSerIleTrpGluLeuGluAspIleLysLysGlyLeuLeuCysGlnLeuPheGlyGly 520
Db    3054 CCAAGCATTTGGGAGCTTGAAGATATTAAAAAGGGTCTTCTTTGCCAGCTCTTTGGTGGG 3113
Qy     521 LysAlaLysLysIleProSerGlyAlaSerPheArgGlyAspIleAsnValLeuLeuVal 540
Db    3114 AAGGCTAAGAAAATTCCATCTGGAGCATCTTTCCGAGGTGACATCAATGTTTTACTTGTT 3173
Qy     541 GlyAspProGlyThrSerLysSerGlnLeuLeuGlnTyrValHisLysIleAlaProArg 560
Db    3174 GGGGACCCTGGTACCAGTAAATCTCAGCTGCTTCAGTATGTGCACAAGATAGCTCCTCGT 3233
Qy     561 GlyIleTyrThrSerGlyArgGlySerSerAlaValGlyLeuThrAlaTyrValThrLys 580
Db    3234 GGAATCTACACTAGTGGGCGAGGAAGTTCGGCGGTTGGGCTGACAGCGTATGTAACGAAG 3293
Qy     581 AspProGluThrArgGluThrValLeuGluSerGlyAlaLeuValLeuSerAspArgGly 600
Db    3294 GATCCAGAAACTCGAGAGACGGTATTGGAGAGCGGAGCTTTGGTTCTTAGTGATCGTGGG 3353
Qy     601 IleCysCysIleAspGluPheAspLysMetSerAspAsnAlaArgSerMetLeuHisGlu 620
Db    3354 ATATGCTGTATCGATGAGTTCGACAAAATGTCTGATAATGCCCGAAGCATGCTTCATGAG 3413
Qy     621 ValMetGluGlnGlnThrValSerValAlaLysGlyGlyIleIleAlaSerLeuAsnAla 640
Db    3414 GTAATGGAGCAACAAACGGTATCTGTAGCCAAAGGGGGTATCATTGCCTCGCTGAACGCT 3473
Qy     641 ArgThrSerValLeuAlaCysAlaAsnProSerGlySerArgTyrAsnAlaArgLeuSer 660
Db    3474 CGGACGTCTGTCCTTGCATGTGCAAATCCTAGTGGGTCCCGATACAATGCGCGCCTTTCT 3533
Qy     661 ValIleAspAsnIleGlnLeuProProThrLeuLeuSerArgPheAspLeuIleTyrLeu 680
Db    3534 GTGATTGATAACATCCAGCTTCCTCCAACTCTACTTTCTAGATTTGATTTAATTTACTTA 3593
Qy     681 MetLeuAspLysProAspGluGlnAsnAspArgArgLeuAlaArgHisLeuValAlaLeu 700
Db    3594 ATGCTCGACAAACCAGACGAGCAAAACGATCGTCGTCTCGCCAGGCATCTCGTGGCTTTA 3653
```

```
Qy        701 HisTyrGluAsnTyrGluValSerLysGlnAspAlaLeuAspLeuGlnThrLeuThrAla 720
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       3654 CACTATGAAAACTATGAAGTTTCAAAGCAGGACGCCTTAGATCTACAAACACTTACCGCG 3713

Qy        721 TyrIleThrTyrAlaArgGlnHisValHisProThrLeuSerAspGluAlaAlaGluAsp 740
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       3714 TATATCACCTATGCTCGTCAGCATGTACATCCTACATTAAGTGATGAAGCTGCTGAAGAT 3773

Qy        741 LeuIleAsnGlyTyrValGluMetArgGlnLysGlyAsnPheProGlySerSerLysLys 760
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       3774 TTGATTAATGGCTATGTTGAGATGCGCCAAAAGGGCAACTTTCCTGGAAGCAGTAAAAAG 3833

Qy        761 ValIleThrAlaThrProArgGlnLeuGluSerMetIleArgIleSerGluAlaLeuAla 780
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       3834 GTGATAACAGCCACACCTCGGCAACTCGAAAGTATGATTCGTATCAGTGAAGCCCTAGCT 3893

Qy        781 ArgMetArgPheSerGluValValGluLysValAspAlaAlaGluAlaValArgLeuLeu 800
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       3894 CGAATGAGATTTTCTGAAGTGGTAGAGAAAGTTGATGCAGCAGAAGCTGTGCGCCTTTTA 3953

Qy        801 AspValAlaLeuGlnGlnSerAlaThrAspHisAlaThrGlyThrIleAspMetAspLeu 820
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       3954 GACGTCGCTTTGCAGCAATCTGCTACTGATCATGCAACAGGTACGATAGACATGGATCTT 4013

Qy        821 IleThrThrGlyValSerAlaSerGluArgIleArgArgAlaAsnLeuLeuAlaAlaLeu 840
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       4014 ATCACGACTGGAGTGTCGGCCAGCGAGCGTATTCGTCGGGCCAACTTGCTAGCTGCTCTG 4073

Qy        841 ArgGluLeuIleAlaAspLysIleSerProGlySerSerSerGlyLeuLysThrSerGln 860
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       4074 CGAGAGCTTATAGCAGATAAAATTTCACCTGGCAGCTCCTCTGGCTTGAAGACCAGTCAG 4133

Qy        861 LeuLeuGluAspIleArgSerGlnSerSerValAspValSerLeuGlnAspIleLysAsn 880
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       4134 CTTCTTGAGGATATCCGGAGCCAAAGCAGTGTGGACGTTAGTTTGCAGGATATTAAAAAT 4193

Qy        881 AlaLeuGlySerLeuGlnGlyGluGlyPheLeuThrValHisGlyAspIleValLysArg 900
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       4194 GCTCTGGGTAGCCTCCAAGGAGAAGGCTTTCTTACTGTCCATGGTGACATAGTCAAGAGA 4253

Qy        901 Val 901
              |||
Db       4254 GTT 4256

RESULT 2
AX281543
LOCUS       AX281543                 4348 bp    DNA     linear   PAT 02-NOV-2001
DEFINITION  Sequence 5 from Patent WO0177354.
ACCESSION   AX281543
VERSION     AX281543.1  GI:16608798
KEYWORDS    .
SOURCE      Physcomitrella patens
  ORGANISM  Physcomitrella patens
            Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Bryophyta;
            Bryopsida; Funariidae; Funariales; Funariaceae; Physcomitrella.
REFERENCE   1
  AUTHORS   costa e Silva,O.D., Bohnert,H.J., van Thielen,N., Chen,R. and
            Sarria-Millan,R.
  TITLE     Cell cycle stress-related proteins and methods of use in plants
  JOURNAL   Patent: WO 0177354-A 5 18-OCT-2001;
            BASF Plant Science GmbH (DE)
FEATURES             Location/Qualifiers
     source          1..4348
                     /organism="Physcomitrella patens"
                     /mol_type="unassigned DNA"
                     /db_xref="taxon:3218"
ORIGIN

Alignment Scores:
Pred. No.:              1.23e-162      Length:         4348
Score:                  4609.00        Matches:        901
Percent Similarity:     100.00%        Conservative:   0
Best Local Similarity:  100.00%        Mismatches:     0

Query Match:            100.00%        Indels:         0
DB:                     6              Gaps:           0

US-10-768-511-8 (1-901) x AX281543 (1-4348)

Qy          1 MetGluAsnAsnAspAlaLeuAspIleGlyAlaValSerSerProTyrProSerGlnSer 20
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1554 ATGGAAAATAATGATGCACTTGACATTGGAGCCGTGTCGTCCCCATATCCTTCGCAATCT 1613

Qy         21 GluGlyValSerThrProLeuProGlnValThrSerProSerPheAspAsnAlaAlaSer 40
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1614 GAAGGAGTGTCTACGCCATTGCCGCAAGTAACATCACCGAGCTTCGACAATGCAGCCTCA 1673

Qy         41 ProValAlaGlyArgArgAlaValArgGlnThrProThrSerAlaValArgArgArgGly 60
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1674 CCCGTGGCCGGGCGGAGGGCCGTACGGCAGACCCCTACATCTGCAGTTCGAAGGAGAGGG 1733

Qy         61 ArgGluThrAspSerAlaArgArgArgArgSerArgSerArgSerLeuGlyAsnSerVal 80
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1734 AGAGAAACGGATTCCGCTCGTCGTAGGAGGAGTCGATCTCGCAGTTTAGGCAATTCTGTT 1793

Qy         81 TyrSerSerProTyrAspAlaGlyThrProGlyThrProGlyThrProValAlaThrPro 100
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1794 TATAGTTCCCCTTACGATGCGGGGACTCCTGGAACTCCTGGAACTCCAGTGGCTACTCCG 1853

Qy        101 ValTyrAlaThrProValGlyThrProMetGlyThrProSerPheHisArgGlyThrPro 120
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1854 GTTTACGCTACCCCAGTCGGTACACCTATGGGTACCCCATCGTTCCATCGTGGCACGCCA 1913

Qy        121 GlnTyrLysGlnArgSerGluLeuGlySerGlnGlyLysProLeuHisArgArgArgArg 140
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1914 CAGTACAAACAGCGCAGTGAGCTTGGTTCCCAGGGGAAGCCTCTACATCGGAGACGTCGA 1973

Qy        141 SerGlnSerArgGluProGlyHisArgSerProSerArgGluProSerAlaAspGlyArg 160
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1974 TCTCAATCCAGAGAACCCGGGCATCGATCTCCTTCAAGGGAACCTAGTGCTGATGGGCGT 2033

Qy        161 ProSerGluSerAlaGluProAspAspThrLeuGlyGlyGluTyrAlaTyrValTrpGly 180
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       2034 CCCTCTGAATCTGCTGAGCCAGATGACACTTTGGGTGGAGAATATGCTTATGTTTGGGGG 2093

Qy        181 ThrAsnValAsnIleProAspValLeuArgAlaIleArgArgPheLeuHisAsnTyrArg 200
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       2094 ACGAATGTTAACATTCCAGATGTGCTTAGGGCGATTCGTCGATTTCTCCACAATTATCGT 2153

Qy        201 SerSerAlaHisAspLeuAsnSerLysTyrIleGlnIleIleGluGluThrValGluArg 220
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       2154 TCGAGTGCTCATGATCTTAATTCCAAGTACATCCAGATCATAGAGGAGACTGTGGAGCGT 2213

Qy        221 GluGluAspThrLeuAsnIleAspMetSerAspIleTyrAspHisAspProAspLeuTyr 240
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       2214 GAGGAGGATACTCTAAATATCGACATGTCAGACATTTATGACCATGATCCTGATCTATAC 2273

Qy        241 AlaLysIleValArgTyrProLeuAspIleIleProLeuLeuAspThrGluCysGlnGlu 260
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       2274 GCAAAAATTGTTCGATACCCACTCGACATCATCCCCCTGTTGGACACTGAGTGTCAGGAA 2333

Qy        261 ValAlaThrSerLeuLeuProThrPheGluLysHisIleGluAlaArgProPheAsnLeu 280
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       2334 GTTGCTACCTCTTTACTACCAACGTTTGAGAAGCATATTGAGGCCAGACCTTTCAATCTC 2393

Qy        281 LysAlaSerValHisMetArgGluLeuAsnProSerAspIleAspLysLeuValSerVal 300
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       2394 AAAGCATCGGTGCACATGCGTGAACTCAACCCTTCAGATATAGACAAATTGGTTTCTGTT 2453

Qy        301 LysGlyMetValIleArgCysSerSerIleIleProGluIleLysGlyAlaPhePheLys 320
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       2454 AAAGGAATGGTTATCCGGTGCAGTTCTATCATACCTGAAATTAAGGGGGCCTTCTTCAAA 2513

Qy        321 CysLeuValCysGlyHisSerProProLeuValThrValValLysGlyArgValGluGlu 340
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       2514 TGTTTAGTGTGTGGTCACTCGCCTCCGCTAGTTACAGTTGTTAAAGGGCGGGTTGAGGAG 2573
```

```
Qy        341 ProThrArgCysGluLysProGluCysAlaAlaArgAsnAlaMetSerLeuIleHisAsn 360
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       2574 CCAACAAGGTGTGAAAAGCCAGAATGTGCAGCACGGAATGCTATGTCTCTTATTCACAAT 2633

Qy        361 ArgCysThrPheAlaAsnLysGlnIleValArgLeuGlnGluThrProAspAlaIlePro 380
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       2634 CGATGCACTTTTGCAAATAAGCAGATAGTGCGTCTTCAAGAAACTCCAGATGCCATTCCT 2693

Qy        381 GluGlyGluThrProHisThrValSerMetCysLeuTyrAsnThrMetValAspAlaVal 400
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       2694 GAAGGAGAGACTCCACACACAGTCAGCATGTGTTTATACAACACTATGGTTGATGCTGTG 2753

Qy        401 LysProGlyAspArgIleGluValThrGlyValPheLysAlaMetAlaValArgValGly 420
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       2754 AAGCCTGGAGATCGTATTGAGGTAACAGGAGTTTTCAAGGCCATGGCAGTTCGAGTTGGT 2813

Qy        421 ProAsnGlnArgThrLeuArgAlaLeuTyrLysThrTyrIleAspCysValHisValLys 440
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       2814 CCGAATCAACGAACATTACGAGCATTGTATAAGACCTACATCGATTGCGTGCACGTCAAG 2873

Qy        441 LysSerAspArgGlyArgLeuGlnThrGluAspProMetGluMetAspLysGluAsnAsp 460
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       2874 AAGTCTGACAGGGGTCGACTGCAAACTGAAGATCCTATGGAGATGGATAAGGAGAATGAT 2933

Qy        461 MetTyrAlaGlyTyrHisGluSerAspThrSerGluAlaAlaAsnGluAlaLysIleGln 480
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       2934 ATGTATGCTGGGTATCATGAAAGTGATACTTCAGAAGCTGCTAATGAAGCAAAGATTCAA 2993

Qy        481 LysLeuLysGluLeuSerLysLeuProGlyIleTyrAspArgLeuSerArgSerLeuAla 500
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       2994 AAACTTAAAGAGCTGTCCAAGCTCCCGGGCATTTATGATAGACTTTCAAGGTCGCTGGCT 3053

Qy        501 ProSerIleTrpGluLeuGluAspIleLysLysGlyLeuLeuCysGlnLeuPheGlyGly 520
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       3054 CCAAGCATTTGGGAGCTTGAAGATATTAAAAAGGGTCTTCTTTGCCAGCTCTTTGGTGGG 3113

Qy        521 LysAlaLysLysIleProSerGlyAlaSerPheArgGlyAspIleAsnValLeuLeuVal 540
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       3114 AAGGCTAAGAAAATTCCATCTGGAGCATCTTTCCGAGGTGACATCAATGTTTTACTTGTT 3173

Qy        541 GlyAspProGlyThrSerLysSerGlnLeuLeuGlnTyrValHisLysIleAlaProArg 560
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       3174 GGGGACCCTGGTACCAGTAAATCTCAGCTGCTTCAGTATGTGCACAAGATAGCTCCTCGT 3233

Qy        561 GlyIleTyrThrSerGlyArgGlySerSerAlaValGlyLeuThrAlaTyrValThrLys 580
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       3234 GGAATCTACACTAGTGGGCGAGGAAGTTCGGCGGTTGGGCTGACAGCGTATGTAACGAAG 3293

Qy        581 AspProGluThrArgGluThrValLeuGluSerGlyAlaLeuValLeuSerAspArgGly 600
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       3294 GATCCAGAAACTCGAGAGACGGTATTGGAGAGCGGAGCTTTGGTTCTTAGTGATCGTGGG 3353

Qy        601 IleCysCysIleAspGluPheAspLysMetSerAspAsnAlaArgSerMetLeuHisGlu 620
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       3354 ATATGCTGTATCGATGAGTTCGACAAAATGTCTGATAATGCCCGAAGCATGCTTCATGAG 3413

Qy        621 ValMetGluGlnGlnThrValSerValAlaLysGlyGlyIleIleAlaSerLeuAsnAla 640
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       3414 GTAATGGAGCAACAAACGGTATCTGTAGCCAAAGGGGGTATCATTGCCTCGCTGAACGCT 3473

Qy        641 ArgThrSerValLeuAlaCysAlaAsnProSerGlySerArgTyrAsnAlaArgLeuSer 660
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       3474 CGGACGTCTGTCCTTGCATGTGCAAATCCTAGTGGGTCCCGATACAATGCGCGCCTTTCT 3533

Qy        661 ValIleAspAsnIleGlnLeuProProThrLeuLeuSerArgPheAspLeuIleTyrLeu 680
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       3534 GTGATTGATAACATCCAGCTTCCTCCAACTCTACTTTCTAGATTTGATTTAATTTACTTA 3593

Qy        681 MetLeuAspLysProAspGluGlnAsnAspArgArgLeuAlaArgHisLeuValAlaLeu 700
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       3594 ATGCTCGACAAACCAGACGAGCAAAACGATCGTCGTCTCGCCAGGCATCTCGTGGCTTTA 3653

Qy        701 HisTyrGluAsnTyrGluValSerLysGlnAspAlaLeuAspLeuGlnThrLeuThrAla 720
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       3654 CACTATGAAAACTATGAAGTTTCAAAGCAGGACGCCTTAGATCTACAAACACTTACCGCG 3713

Qy        721 TyrIleThrTyrAlaArgGlnHisValHisProThrLeuSerAspGluAlaAlaGluAsp 740
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       3714 TATATCACCTATGCTCGTCAGCATGTACATCCTACATTAAGTGATGAAGCTGCTGAAGAT 3773

Qy        741 LeuIleAsnGlyTyrValGluMetArgGlnLysGlyAsnPheProGlySerSerLysLys 760
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       3774 TTGATTAATGGCTATGTTGAGATGCGCCAAAAGGGCAACTTTCCTGGAAGCAGTAAAAAG 3833

Qy        761 ValIleThrAlaThrProArgGlnLeuGluSerMetIleArgIleSerGluAlaLeuAla 780
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       3834 GTGATAACAGCCACACCTCGGCAACTCGAAAGTATGATTCGTATCAGTGAAGCCCTAGCT 3893

Qy        781 ArgMetArgPheSerGluValValGluLysValAspAlaAlaGluAlaValArgLeuLeu 800
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       3894 CGAATGAGATTTTCTGAAGTGGTAGAGAAAGTTGATGCAGCAGAAGCTGTGCGCCTTTTA 3953

Qy        801 AspValAlaLeuGlnGlnSerAlaThrAspHisAlaThrGlyThrIleAspMetAspLeu 820
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       3954 GACGTCGCTTTGCAGCAATCTGCTACTGATCATGCAACAGGTACGATAGACATGGATCTT 4013

Qy        821 IleThrThrGlyValSerAlaSerGluArgIleArgArgAlaAsnLeuLeuAlaAlaLeu 840
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       4014 ATCACGACTGGAGTGTCGGCCAGCGAGCGTATTCGTCGGGCCAACTTGCTAGCTGCTCTG 4073

Qy        841 ArgGluLeuIleAlaAspLysIleSerProGlySerSerSerGlyLeuLysThrSerGln 860
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       4074 CGAGAGCTTATAGCAGATAAAATTTCACCTGGCAGCTCCTCTGGCTTGAAGACCAGTCAG 4133

Qy        861 LeuLeuGluAspIleArgSerGlnSerSerValAspValSerLeuGlnAspIleLysAsn 880
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       4134 CTTCTTGAGGATATCCGGAGCCAAAGCAGTGTGGACGTTAGTTTGCAGGATATTAAAAAT 4193

Qy        881 AlaLeuGlySerLeuGlnGlyGluGlyPheLeuThrValHisGlyAspIleValLysArg 900
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       4194 GCTCTGGGTAGCCTCCAAGGAGAAGGCTTTCTTACTGTCCATGGTGACATAGTCAAGAGA 4253

Qy        901 Val 901
              |||
Db       4254 GTT 4256
```

RESULT 3
CR954214_1/c
WPCOMMENT
Sequence split into 6 fragments LOCUS CR954214 Accession CR954214

Fragment Name	Begin	End
CR954214_0	1	110000
CR954214_1	100001	210000
CR954214_2	200001	310000
CR954214_3	300001	410000
CR954214_4	400001	510000
CR954214_5	500001	516206

Continuation (2 of 6) of CR954214 from base 100001 (CR954214 Ostreococcus tauri strain

Alignment Scores:

Pred. No.:	6.38e-65	Length:	110000
Score:	2016.00	Matches:	446
Percent Similarity:	60.30%	Conservative:	154
Best Local Similarity:	44.82%	Mismatches:	229
Query Match:	43.74%	Indels:	166
DB:	14	Gaps:	25

US-10-768-511-8 (1-901) x CR954214_1 (1-110000)

```
Qy         39 AlaSerProValAlaGlyArg---ArgAlaValArgGlnThrProThrSerAlaValArg 57
              |||||||||       |||     |||:::       :::::|||:::||||||
Db      73209 GCGTCGCCCGCTCGCGCGCTCGCGGTCACCCGTGAAGTCACCGTCGTCCGCG---TCG 73153

Qy         58 ArgArgGlyArgGluThr------------------AspSerAlaArgArgArgArgSer 71
              :::|||||||    |||                          ||||||:::
Db      73152 CAGCGCGGGCGCGCGACTTTACACGACGCGAACGCGGCGGACGGTGAGGGGCGACGAGCG 73093
```

```
Qy         72 ArgSerArgSerLeuGlyAsnSer------------------------------------ 79
Db      73092 CTCGATCGATCGATGGGCGAAGCGACAGCTGTGGATGAAAACGTCGCCCCGGACGGGCGA 73033
Qy         80 ------------------------------------ValTyrSerSerProTyrAspAla 87
Db      73032 GGGACGAGTGCGGAGGGGGCTGGGGGCGATGTGGGGTTGATGTCGCCGCCGCCGGCGCGC 72973
Qy         88 GlyThrProGlyThrProGly------------ThrProValAlaThrProValTyrAla 103
Db      72972 GGAGCGAGCTCGGGGGGTGGACGAGGGATGATGTCGCCGTTGGATGCGGATTTGTTTTCC 72913
Qy        104 ThrPro---------------------ValGlyThrProMet---------------Gly 111
Db      72912 CCGCCGGCGAGACGCGCGCACGGCGCGGCGTCCACGCCGATGGGTGGTGGGACGTTTGGA 72853
Qy        112 ThrProSerPheHisArgGlyThrProGlnTyrLysGlnArgSerGluLeuGlySerGln 131
Db      72852 TCACCGCGGTTC------GGGACGCCGAGGACGACGCCGAGGGGGGATGTT---CGACGA 72802
Qy        132 GlyLysProLeuHisArgArgArgArgSerGlnSerArgGluProGlyHisArgSerPro 151
Db      72801 GGGGTGTTGAGTGCACGCGCGGGACGGATGGATGAGGAGGACGATGGATTCGGCGCGCCG 72742
Qy        152 SerArgGluProSerAlaAspGly------------------------------------ 159
Db      72741 GGGAGCGAGCGCCGGGCGAGCGCGGGCTTGGACGCGACTCCTGGAACGCAGTTGACGGCG 72682
Qy        160 ------------------------ArgProSerGluSerAlaGluProAspAspThrLeu 171
Db      72681 ACGCAGGATGTTGATTTCTTGATGTCACCGCCGCAGAGCGCG---GGTGATCACTTGATG 72625
Qy        172 GlyGlyGlu------------------------TyrAlaTyrValTrpGlyThrAsnVal 183
Db      72624 CGTGGGCACATGGATCGCGAGGACGTGGCGATGCAGACGTATATCTGGGGTACCAAGGTG 72565
Qy        184 AsnIleProAspValLeuArgAlaIleArgArgPheLeuHisAsnTyrArg--------- 200
Db      72564 AACGTGTACGACGTGCAGACGCGCTTTCGGCGATTTTTGGAAAACTTTGAGTGGTATGTC 72505
Qy        200 ------------------------------------------------------------ 200
Db      72504 CTTCCGTCGGAGTTTCAGTGGCCTGAGGGCTACGGTGAGGCTGATTTTGTGCCTCCGGTT 72445
Qy        201 ---------------------SerSerAlaHisAspLeuAsnSerLysTyrIleGlnIle 213
Db      72444 GATTTCGTTCCGCCACCGACTCGTAACTACCACGACTTG--------------------- 72406
Qy        214 IleGluGluThrValGluArgGluGluAspThrLeuAsnIleAspMetSerAspIleTyr 233
Db      72405 ATGCACGATCTCCTGGAGAAGCAGCAATCGGATCTGGATCTGGATTGTCAGCACGTCGAC 72346
Qy        234 AspHisAspProAspLeuTyrAlaLysIleValArgTyrProLeuAspIleIleProLeu 253
Db      72345 AAATACGATCCGTTCTTGTACAGTCTCTTGACGCAGTACCCGCAAGAGATCATTCCGCTC 72286
Qy        254 LeuAspThrGluCys---------------GlnGluValAlaThrSerLeuLeuProThr 268
Db      72285 TTCGACGTTGTTGCAAATGAATACTTCACGCAAAATGTCGTGCCAGAAGTGATGCCGGTT 72226
Qy        269 PheGlu---------LysHisIleGluAlaArgProPheAsnLeuLysAlaSerValHis 285
Db      72225 GATGAGGCCGATGATCACCGATTGCTAGTGCGCCCGTTCAACATGATGGAGGCGAAACCC 72166
Qy        286 MetArgGluLeuAsnProSerAspIleAspLysLeuValSerValLysGlyMetValIle 305
Db      72165 ATGCGCGACTTGAATCCATCGGACATCGACAAGATGGTGTGCGTGCGTGGTATGGTGACG 72106
Qy        306 ArgCysSerSerIleIleProGluIleLysGlyAlaPhePheLysCysLeuValCysGly 325
Db      72105 CGATGCACAACGATCATTCCTGACCTAAAACTCGCGTATTTCAAGTGTCTGATGTGCGGC 72046
Qy        326 HisSerProProLeuValThrValValLysGlyArgValGluGluPro------------ 341
Db      72045 TTCGCCCCAGAACATATTCAAGTCGACCGCGGTCGCGTCAACGAACCTCCGCTCAAGTGC 71986
Qy        342 ThrArgCysGluLysProGluCysAlaAlaArgAsnAlaMetSerLeuIleHisAsnArg 361
Db      71985 ACAGAGTGCGGCAAGCCA---------------GGAACGATGACCTTGATTCACAATCAA 71941
Qy        362 CysThrPheAlaAsnLysGlnIleValArgLeuGlnGluThrProAspAlaIleProGlu 381
Db      71940 TGCGTCTTCGCAAACAAGCAAACGGTGAAGATGCAGGAGACACCGGATGCGATTCCCGAG 71881
Qy        382 GlyGluThrProHisThrValSerMetCysLeuTyrAsnThrMetValAspAlaValLys 401
Db      71880 GGTGAGACTCCGCACACTGTATCCATGTGCGTCTTCGATGATCTCGTTGATCAGGCGAAG 71821
Qy        402 ProGlyAspArgIleGluValThrGlyValPheLysAlaMetAlaValArgValGlyPro 421
Db      71820 CCAGGTGACCGAGTTGAAGTTACTGGAGTGTATCGTGCGGTTCCAATTCGTCTGAGCTCT 71761
Qy        422 AsnGlnArgThrLeuArgAlaLeuTyrLysThrTyrIleAspCysValHisValLysLys 441
Db      71760 ACGAAGCGCACGCTCAAGAGTGTGTACAAGACGTACCTAGACGTTTTTCATATAAGAAAG 71701
Qy        442 SerAspArgGlyArgLeuGln---ThrGluAspProMetGluMetAspLysGluAsnAsp 460
Db      71700 GACGTCGGGGCGAGGATGCGAAACACGGCCGGACCC------------------GAGGAT 71659
Qy        461 MetTyrAlaGlyTyrHisGluSerAspThrSerGluAlaAlaAsnGluAlaLysIleGln 480
Db      71658 GAAGAGGCTGCAAGAAACTCAGCTGCTTCAACGAAGTCATCTGGACCGGTGAAGAATCAA 71599
Qy        481 ------------------------------------------LysLeuLysGluLeuSer 486
Db      71598 GGTCCTGGACAGCAAATGGAGTTCACTCCGGAGCGCATGGCCGAGATAGAAGAGCTCGGT 71539
Qy        487 LysLeuProGlyIleTyrAspArgLeuSerArgSerLeuAlaProSerIleTrpGluLeu 506
Db      71538 CGTTCGCCGGACATTTATGACCGGCTTGTGGCGTCGTTGGCCCCGTCAATTTGGGAGTTA 71479
Qy        507 GluAspIleLysLysGlyLeuLeuCysGlnLeuPheGlyGlyLysAlaLysLysIlePro 526
Db      71478 GAAGACGTTAAAAAAGGACTGTTGTGTCAGCTCTTTGGCGCCACGAACAAGAGCTTCAGC 71419
Qy        527 SerGlyAlaSer------PheArgGlyAspIleAsnValLeuLeuValGlyAspProGly 544
Db      71418 GACAAGGCAGCGAACAAGGTCCGGGGCGACATCAACATTTTACTCGTCGGTGATCCCGGG 71359
Qy        545 ThrSerLysSerGlnLeuLeuGlnTyrValHisLysIleAlaProArgGlyIleTyrThr 564
Db      71358 GTCGCCAAGTCTCAGCTATTGACGTACGTTCATCGTATCGCGCCGCGTGGTATGTACACC 71299
Qy        565 SerGlyArgGlySerSerAlaValGlyLeuThrAlaTyrValThrLysAspProGluThr 584
Db      71298 TCAGGTAGAGGCTCGTCTGCGGTTGGTCTTACAGCGTACGTCACTCGCGATCCGGAGTCA 71239
Qy        585 ArgGluThrValLeuGluSerGlyAlaLeuValLeuSerAspArgGlyIleCysCysIle 604
Db      71238 AAAGACATGGTGCTCGAGTCAGGAGCGCTCGTGCTTTCTGATCGCGGCATTTGTTGCATT 71179
Qy        605 AspGluPheAspLysMetSerAspAsnAlaArgSerMetLeuHisGluValMetGluGln 624
Db      71178 GATGAGTTCGACAAGATGTCTGACAGTGCTCGATCAATGTTGCACGAAGTGATGGAGCAG 71119
Qy        625 GlnThrValSerValAlaLysGlyGlyIleIleAlaSerLeuAsnAlaArgThrSerVal 644
Db      71118 CAGACGGTTTCCATCGCGAAGGCTGGAATCATTGCGGTGTTAAATGCACGGACTTCGGTC 71059
Qy        645 LeuAlaCysAlaAsnProSerGlySerArgTyrAsnAlaArgLeuSerValIleAspAsn 664
Db      71058 CTTGCGTCAGCGAACCCAGTCGGAAGTCGTTACAACCCGAACATGTCCATGGTGGAGAAC 70999
Qy        665 IleGlnLeuProProThrLeuLeuSerArgPheAspLeuIleTyrLeuMetLeuAspLys 684
```

```
                ||||||||||||||||||||||||||||||||||||||:::||||||:::||||||:::
Db      70998 ATTCAACTCCCGCCGACACTGTTGTCCCGTTTTGATTTGCTCTACCTATTGCTTGACCGT 70939

Qy        685 ProAspGluGlnAsnAspArgArgLeuAlaArgHisLeuValAlaLeuHisTyrGluAsn 704
              |||:::   :::   |||||||||||||||||||||||||:::|||||||||:::|||
Db      70938 CCTAATCCCGAGACGGATCGCCGTCTGGCGAGACACTTAGTCAGTTTGCACTACAAGAAT 70879

Qy        705 TyrGluValSerLysGlnAspAlaLeuAspLeuGlnThrLeuThrAlaTyrIleThrTyr 724
                             |||:::      :::            ||||||   |||::::::|||
Db      70878 CCCCCGCAGAAGAAACGGGGCGTGATTTCTGCCGACTTGCTCACCGAGTACGTCTCGTAT 70819

Qy        725 AlaArgGlnHisValHisProThrLeuSerAspGluAlaAlaGluAspLeuIleAsnGly 744
              ||||||   :::|||   |||   |||||||||||||||:::|||:::|||:::   |||
Db      70818 GCTCGCGCCAATGTACAACCGGTGCTCAGCGACGAAGCCTCGGAAGAGCTCGTGGAGGGA 70759

Qy        745 TyrValGluMetArgGlnLysGlyAsnPheProGlySerSerLysLysValIleThrAla 764
              ||||||||||||||:::                    |||    |||:::||||||||||||||
Db      70758 TATGTTGAGATGCGACGCATG------------GGTGGTAGTCGAAAAGTTATTACTGCA 70711

Qy        765 ThrProArgGlnLeuGluSerMetIleArgIleSerGluAlaLeuAlaArgMetArgPhe 784
              ||||||||||||||||||||:::||||||:::||||||:::||||||||||||||
Db      70710 ACGCCGAGACAACTTGAGTCTCTCATTCGTCTCTCAGAGTCTCTAGCGCGCATGCGTTTA 70651

Qy        785 SerGluValValGluLysValAspAlaAlaGluAlaValArgLeuLeuAspValAlaLeu 804
              |||   ||||||::::::   ||||||   ||||||:::|||||||:::   ||||||:::
Db      70650 AGTGCGGTGGTGGATCGTGACGACGCTAAGGAGGCGCTGCGATTGATGCGGGTGGCGATG 70591

Qy        805 GlnGlnSerAlaThrAspHisAlaThrGlyThrIleAspMetAspLeuIleThrThrGly 824
              ||||||||||||   |||      ||||||||||||||||||||||   |||   ||||||
Db      70590 CAACAATCCGCCGTTGATCCTCGCACTGGGACGATTGATATGGATAAGATATTGACCGGA 70531

Qy        825 ValSerAlaSerGluArgIleArgArgAlaAsnLeuLeuAlaAlaLeuArgGluLeuIle 844
                 |||||||||:::|||      |||      :::      |||:::|||
Db      70530 CATAGCGCCAGCGACCGTCAGCACCGCAGGACGGTTGCAGAGGCCATCAGGGCGTGT--- 70474

Qy        845 AlaAspLysIleSerProGlySerSerSerGlyLeuLysThrSerGlnLeuLeuGluAsp 864
                       ::::::         |||      |||:::      :::|||   :::
Db      70473 ---------CTTAATACTTCTGGGAGCGGTAGACTCCGACTGGGCGAGCTCGCGAAGGCG 70423

Qy        865 IleArgSerGln---SerSerValAspValSerLeuGlnAspIleLysAsnAlaLeuGly 883
              :::      :::   :::::::::   :::|||:::|||::::::::::::|||
Db      70422 CTCGCGGAGAGGGACAACACGATGGCGCTGAGCATCCAGGAGGTTCGCGACGCCGCAATG 70363

Qy        884 SerLeuGlnGlyGluGlyPheLeuThrValHisGlyAspIleVal 898
                 |||      |||         |||:::   ||||||:::|||
Db      70362 CTTCTTGTAGAGGAGGAAAAGTGTACCATTCGAGGTGACCTCGTC 70318
```

RESULT 4
AP008207_201
WPCOMMENT
Sequence split into 433 fragments LOCUS AP008207 Accession AP008207

Fragment Name	Begin	End
AP008207_000	1	110000
AP008207_001	100001	210000
AP008207_002	200001	310000
AP008207_003	300001	410000
AP008207_004	400001	510000
AP008207_005	500001	610000
AP008207_006	600001	710000
AP008207_007	700001	810000
AP008207_008	800001	910000
AP008207_009	900001	1010000
AP008207_010	1000001	1110000
AP008207_011	1100001	1210000
AP008207_012	1200001	1310000
AP008207_013	1300001	1410000
AP008207_014	1400001	1510000
AP008207_015	1500001	1610000
AP008207_016	1600001	1710000
AP008207_017	1700001	1810000
AP008207_018	1800001	1910000
AP008207_019	1900001	2010000
AP008207_020	2000001	2110000
AP008207_021	2100001	2210000
AP008207_022	2200001	2310000
AP008207_023	2300001	2410000
AP008207_024	2400001	2510000
AP008207_025	2500001	2610000
AP008207_026	2600001	2710000
AP008207_027	2700001	2810000
AP008207_028	2800001	2910000
AP008207_029	2900001	3010000
AP008207_030	3000001	3110000
AP008207_031	3100001	3210000
AP008207_032	3200001	3310000
AP008207_033	3300001	3410000
AP008207_034	3400001	3510000
AP008207_035	3500001	3610000
AP008207_036	3600001	3710000
AP008207_037	3700001	3810000
AP008207_038	3800001	3910000
AP008207_039	3900001	4010000
AP008207_040	4000001	4110000
AP008207_041	4100001	4210000
AP008207_042	4200001	4310000
AP008207_043	4300001	4410000
AP008207_044	4400001	4510000
AP008207_045	4500001	4610000
AP008207_046	4600001	4710000
AP008207_047	4700001	4810000
AP008207_048	4800001	4910000
AP008207_049	4900001	5010000
AP008207_050	5000001	5110000
AP008207_051	5100001	5210000
AP008207_052	5200001	5310000
AP008207_053	5300001	5410000
AP008207_054	5400001	5510000
AP008207_055	5500001	5610000
AP008207_056	5600001	5710000
AP008207_057	5700001	5810000
AP008207_058	5800001	5910000
AP008207_059	5900001	6010000
AP008207_060	6000001	6110000
AP008207_061	6100001	6210000
AP008207_062	6200001	6310000
AP008207_063	6300001	6410000
AP008207_064	6400001	6510000
AP008207_065	6500001	6610000
AP008207_066	6600001	6710000
AP008207_067	6700001	6810000
AP008207_068	6800001	6910000
AP008207_069	6900001	7010000
AP008207_070	7000001	7110000
AP008207_071	7100001	7210000
AP008207_072	7200001	7310000
AP008207_073	7300001	7410000
AP008207_074	7400001	7510000
AP008207_075	7500001	7610000
AP008207_076	7600001	7710000
AP008207_077	7700001	7810000
AP008207_078	7800001	7910000
AP008207_079	7900001	8010000
AP008207_080	8000001	8110000
AP008207_081	8100001	8210000
AP008207_082	8200001	8310000
AP008207_083	8300001	8410000
AP008207_084	8400001	8510000
AP008207_085	8500001	8610000
AP008207_086	8600001	8710000
AP008207_087	8700001	8810000
AP008207_088	8800001	8910000
AP008207_089	8900001	9010000
AP008207_090	9000001	9110000
AP008207_091	9100001	9210000
AP008207_092	9200001	9310000
AP008207_093	9300001	9410000

```
AP008207_094      9400001    9510000
AP008207_095      9500001    9610000
AP008207_096      9600001    9710000
AP008207_097      9700001    9810000
AP008207_098      9800001    9910000
AP008207_099      9900001    10010000
AP008207_100      10000001   10110000
AP008207_101      10100001   10210000
AP008207_102      10200001   10310000
AP008207_103      10300001   10410000
AP008207_104      10400001   10510000
AP008207_105      10500001   10610000
AP008207_106      10600001   10710000
AP008207_107      10700001   10810000
AP008207_108      10800001   10910000
AP008207_109      10900001   11010000
AP008207_110      11000001   11110000
AP008207_111      11100001   11210000
AP008207_112      11200001   11310000
AP008207_113      11300001   11410000
AP008207_114      11400001   11510000
AP008207_115      11500001   11610000
AP008207_116      11600001   11710000
AP008207_117      11700001   11810000
AP008207_118      11800001   11910000
AP008207_119      11900001   12010000
AP008207_120      12000001   12110000
AP008207_121      12100001   12210000
AP008207_122      12200001   12310000
AP008207_123      12300001   12410000
AP008207_124      12400001   12510000
AP008207_125      12500001   12610000
AP008207_126      12600001   12710000
AP008207_127      12700001   12810000
AP008207_128      12800001   12910000
AP008207_129      12900001   13010000
AP008207_130      13000001   13110000
AP008207_131      13100001   13210000
AP008207_132      13200001   13310000
AP008207_133      13300001   13410000
AP008207_134      13400001   13510000
AP008207_135      13500001   13610000
AP008207_136      13600001   13710000
AP008207_137      13700001   13810000
AP008207_138      13800001   13910000
AP008207_139      13900001   14010000
AP008207_140      14000001   14110000
AP008207_141      14100001   14210000
AP008207_142      14200001   14310000
AP008207_143      14300001   14410000
AP008207_144      14400001   14510000
AP008207_145      14500001   14610000
AP008207_146      14600001   14710000
AP008207_147      14700001   14810000
AP008207_148      14800001   14910000
AP008207_149      14900001   15010000
AP008207_150      15000001   15110000
AP008207_151      15100001   15210000
AP008207_152      15200001   15310000
AP008207_153      15300001   15410000
AP008207_154      15400001   15510000
AP008207_155      15500001   15610000
AP008207_156      15600001   15710000
AP008207_157      15700001   15810000
AP008207_158      15800001   15910000
AP008207_159      15900001   16010000
AP008207_160      16000001   16110000
AP008207_161      16100001   16210000
AP008207_162      16200001   16310000
AP008207_163      16300001   16410000
AP008207_164      16400001   16510000
AP008207_165      16500001   16610000
AP008207_166      16600001   16710000
AP008207_167      16700001   16810000
AP008207_168      16800001   16910000
AP008207_169      16900001   17010000
AP008207_170      17000001   17110000
AP008207_171      17100001   17210000
AP008207_172      17200001   17310000
AP008207_173      17300001   17410000
AP008207_174      17400001   17510000
AP008207_175      17500001   17610000
AP008207_176      17600001   17710000
AP008207_177      17700001   17810000
AP008207_178      17800001   17910000
AP008207_179      17900001   18010000
AP008207_180      18000001   18110000
AP008207_181      18100001   18210000
AP008207_182      18200001   18310000
AP008207_183      18300001   18410000
AP008207_184      18400001   18510000
AP008207_185      18500001   18610000
AP008207_186      18600001   18710000
AP008207_187      18700001   18810000
AP008207_188      18800001   18910000
AP008207_189      18900001   19010000
AP008207_190      19000001   19110000
AP008207_191      19100001   19210000
AP008207_192      19200001   19310000
AP008207_193      19300001   19410000
AP008207_194      19400001   19510000
AP008207_195      19500001   19610000
AP008207_196      19600001   19710000
```

```
Alignment Scores:
Pred. No.:               5.26e-62     Length:         110000
Score:                   1937.50      Matches:        530
Percent Similarity:      41.94%       Conservative:   133
Best Local Similarity:   33.52%       Mismatches:     173
Query Match:             42.04%       Indels:         751
DB:                      15           Gaps:           31

US-10-768-511-8 (1-901) x AP008207_201 (1-110000)

Qy        11 AlaValSerSerProTyrProSerGlnSerGluGlyValSerThrProLeuProGlnVal 30
             ||||||||||||||                             |||:::||||||||
Db     88290 GCCGTGTCGTCGCCG-----------GATGTGCGACCGTCGAGCCCTCTCCCGGCGACC 88337

Qy        31 ThrSerProSerPheAspAsnAlaAlaSerProValAlaGlyArgArgAlaValArgGln 50
                               |||::::::    |||    :::|||||||
Db     88338 -----------------AACTCCTCCCCGCCTCAGTCGGGGCGGCGCGGC--------- 88370

Qy        51 ThrProThrSerAlaValArgArgArgGlyArgGluThrAspSerAlaArgArgArgArg 70
                                                             |||||||||||
Db     88371 ----------------------------------------GGGGGGCGGCGCCGGCGT 88388

Qy        71 SerArgSerArgSerLeuGlyAsnSerValTyrSerSerProTyrAspAlaGlyThrPro 90
                                           :::|||||||||   :::    :::|||
Db     88389 GGTTCC-------------------------GCCAGCCCGTACCCGTCG---TCCCCG 88418

Qy        91 GlyThrProGlyThrProValAlaThrProValTyrAlaThrProValGlyThrProMet 110
                      |||                            :::   ||||||
Db     88419 TCCCTCGGCGGG--------------------TTCGAGACGCCGCCGCACCCGGGCCGC 88457

Qy       111 GlyThrProSerPheHisArgGlyThrProGlnTyrLysGlnArgSerGluLeuGlySer 130
                |||||||||          |||             :::||||||    ||||||
Db     88458 CGCACGCCCTCC------GGCGGCGCGGCGGCGCGGCAGCAGCGGCA-GAACTGGACGGG 88510

Qy       131 GlnGlyLysPro-------------------LeuHisArgArgArgArgSerGlnSer 143
                      |||                     :::|||||||||||||
Db     88511 GGGGCGGTTCCCGCCGACCCCGTCCACCCCCATGTCCACCGACGACGTCC-CGCTCTCCT 88569

Qy       144 ArgGluProGlyHisArgSerProSerArgGluProSerAla--AspGlyArgProSerG 163
             |||    ||||||   |||:::|||    |||       :::|||      |||
Db     88570 CGGAGGCCGGGGACGAGGACACCCCCGAGACCGACGGCGGCGGCGGCGGCGGCGCGGGCG 88629
```

```
Qy        163 luSerAlaGluProAspAspThrLeuGlyGlyGluTyrAlaTyrValTrpGlyThrAsnV 183
Db      88630 CCGACGCCACGCCG----------------------GTGTTCGTCTGGGGCACCAACA 88665

Qy        183 alAsnIleProAspValLeuArgAlaIleArgArgPheLeuHisAsnTyrArgSerSerA 203
Db      88666 TCAGCGTGCAGGACGTCAACGCCGCCATCCTCCGGTTCCTCCGCCACTTCCGC---GACC 88722

Qy        203 laHisAspLeu----------------------AsnSerLysTyrIleGlnIleIleG 215
Db      88723 CGCGCGACGCCGGCCGCGTCGACCCGGTCATGGACGAGGGCAAGTACATGCGCGCCATCC 88782

Qy        215 luGluThrValGluArgGlu---GluAspThrLeuAsnIleAspMetSerAspIleTyrA 234
Db      88783 ACCGCATCCTCGAGCTCGAGGGCGGGGAGTCGCTCGACGTCAACGCCCACGACGTGTTCG 88842

Qy        234 spHisAspProAspLeuTyrAlaLysIleValArgTyrProLeuAspIleIleProLeuL 254
Db      88843 ACCACGACCCGGACCTCTACGGCAAGATGGTCAGGTACCCGCTCGAGGTGCTCGCCATCT 88902

Qy        254 euAspThrGluCysGlnGluValAlaThrSerLeuLeuProThrPheGluLysHisIleG 274
Db      88903 TTGACATCGTGCTCATGGACCTCGTCGCGCGCATCGAGCCCCTCTTCGAGAAGCACATCC 88962

Qy        274 luAlaArgProPheAsnLeuLysAlaSerValHisMetArgGluLeuAsnProSer---- 292
Db      88963 AGACCAGGATCTACAACCTCAAGTCCTCGGTTTGCTTGAGGAATCTCAACCCTTC-TGGT 89021

Qy        292 ------------------------------------------------------------ 292
Db      89022 GATTGCGCACTGCTCTGCTCCTTTAGCTACTTGTGGTCGGCAATTCTTTTATATGATTTC 89081

Qy        293 -------------------AspIleAspLysLeuValSerValLysGlyMetValIleA 306
Db      89082 TGATTTTGCGTTTCGTTGCAGATATCGAGAAGATGGTGTCCATCAAGGGTATGATAATTC 89141

Qy        306 rgCysSerSerIleIleProGluIleLysGlyAlaPhePheLysCysLeuValCysGly- 325
Db      89142 GGTGCAGCTCGGTGATTCCAGAGCTCAAGGAGGCTGTGTTCCGCTGCTTGGTTTGCGGGT 89201

Qy        326 --HisSerProProLeuVal---------------------------------------- 331
Db      89202 TCTACTCTGAGCCTGTAATGGTTGATCGAGGTATATATATATCAGAACTAATGTGGATAC 89261

Qy        331 ------------------------------------------------------------ 331
Db      89262 CTTTTTCACACTGCGTATTCCATATGCTCTTGTTTTAAGTGAACTTCAGGTGATCAATTA 89321

Qy        331 ------------------------------------------------------------ 331
Db      89322 ATCAGATTGCGAGCCTTAATTTTGCTCATAGAGAGTTACATATTGACAACTCATTGTTAA 89381

Qy        331 ------------------------------------------------------------ 331
Db      89382 CTTCATTTGTACAAATCATTCTTATCATGAAAGATGATAAATTCAAGCTTCATTTGACAT 89441

Qy        331 ------------------------------------------------------------ 331
Db      89442 GTTATATGTCATTACTATTATCATTTATGCATTTTCTTTTATTTTGTTTGCAAAAACTTT 89501

Qy        331 ------------------------------------------------------------ 331
Db      89502 TCATTTTGTTGGGATACACAACATTTAGGAATTTTATTTATCATACCACATTGCTGAGGA 89561

Qy        331 ------------------------------------------------------------ 331
Db      89562 ACTGATAAATAGTGATACTAATTGAGTACGCAAATCAAACATCCATACTGCTATTGTCCA 89621

Qy        331 ------------------------------------------------------------ 331
Db      89622 ATCCAAGACTCCGACTTAGTCTGTCTGTCATAAAGTAACACTTTGCACTGTGGTAGTAAT 89681

Qy        331 ------------------------------------------------------------ 331
Db      89682 TGTAAGTAAAAACATGCTTATGTTCTTAACTTACATTATATTCTGGTGGTTAATGTTTGA 89741

Qy        331 ------------------------------------------------------------ 331
Db      89742 AGAGATTGGTCATGCAGATCCTATGAAAACACATTAGTGGCCTAGTGGGAGTAATGACTA 89801

Qy        332 ----------------------------------------ThrVal-ValLysGlyArg 337
Db      89802 AAACTCCATCAGTTTCTGTTGTTTTATCATGTAACTAATTCACTCTATTTCTTAGGGAGG 89861

Qy        338 ValGluGluProThrArgCysGluLysProGluCysAlaAlaArgAsnAlaMetSerLeu 357
Db      89862 GTTACTGAGCCGCACATCTGTCAGAAAGAACAATGTAAAGCTACAAACTCTATGACTCTT 89921

Qy        358 IleHisAsnArgCysThr------------------------------------------ 363
Db      89922 GTGCATAACCGATGCAG-GTAGCTATATCCTACTTGTAACCCCTTTTTCTTTATAGGATT 89980

Qy        364 ------------------------------------------------------PheAla 365
Db      89981 TTTGACAGATATGTAGTCCTGACCTTTGGATGGAACATTTTATTTGTAACTAGGTTTGCA 90040

Qy        366 AsnLysGlnIleValArgLeuGlnGluThrProAspAlaIleProGluGlyGluThrPro 385
Db      90041 GATAAGCAGATCATAAAGTTGCAGGAAACACCAGATGAGATACCAGAAGGTGGCACTCCA 90100

Qy        386 HisThrValSerMetCysLeuTyrAsnThrMetValAspAlaValLysProGlyAspArg 405
Db      90101 CATACTGTTAGTGTCTTGATGCACGATAAGCTTGTAGATGCTGGAAAGCCTGGAGACAGG 90160

Qy        406 IleGluVal--------------------------------------------------- 408
Db      90161 GTTGAGGTTAGCAGTGAATCTTTGTTTGGTCATTAATACAAGAAAATAATTTGTTGTG 90220

Qy        409 ---------------------------------------------------ThrGlyVal 411
Db      90221 GTTCTTATTCACATCGTAAAAATCTTGATTTTTTTCTTTGATTTATAGATAACTGGGATA 90280

Qy        412 PheLysAlaMetAlaValArgValGlyProAsnGlnArgThrLeuArg------------ 427
Db      90281 TACAGAGCCATGAGTATTAGAGTTGGCCCAACTCAGAGGACAGTGAAGTCGATTTTCAAG 90340

Qy        427 ------------------------------------------------------------ 427
Db      90341 GTAGAGGATGTGCAATGTTATCCTACTTGCTATCTTAAATTCCTTTTCTGCACAGGAAAA 90400

Qy        428 ---------------------------------------AlaLeuTyrLysThrTyrIle 434
Db      90401 AAGTTTTAAATTATCTACTCAAGCATCGTCTAACAGTTTTCTTTAAT-CAGACGTACATT 90459

Qy        435 AspCysValHisValLysLysSerAspArgGlyArgLeuGlnThrGluAspProMetGlu 454
Db      90460 GATTGCCTTCACATAAAGAAGACAGACAAGTCTAGACTTCATGTTGAGGACTCCATGGAA 90519

Qy        455 MetAspLysGluAsn---------------------------------AspMetTyrAla 463
Db      90520 ACTGATAACCCCAATGCTAACAAGACAACTGAAGATGATTTTCTCAGAGATAAGGTATCC 90579

Qy        464 GlyTyrHisGluSerAspThrSerGluAla--------------------------Ala 474
Db      90580 CTCAATCACCTTACCTCTCTGGGATA-GCACTAAGTGTAAATGTGTTTAACACTGTCCTT 90638

Qy        475 AsnGluAlaLysIleGlnLysLeuLysGluLeuSerLysLeuProGlyIleTyrAspArg 494
Db      90639 TCTGTATTGCAGGTTGAGAAATTAAAAGAGTTGTCAAAGTTGCCAGATATATATGACAGA 90698

Qy        495 LeuSerArgSerLeuAlaProSerIleTrpGluLeuGluAspIleLysLysGly------ 512
Db      90699 TTAACTAGGTCATTGGCTCCAAACATATGGGAGCTGGACGACGTTAAAAGGGGCCTCCTT 90758

Qy        512 ------------------------------------------------------------ 512
```

```
Db  90759 TGCCAGGTAAGATCATCTATCTGTTTGAATGTTTTTGCACTAATAGCCTAGTATAACAAG 90818

Qy    513 ------LeuLeuCys-GlnLeuPheGlyGlyLysAlaLysLysIleProSerGlyAlaSe 530
Db  90819 CTCTTACTGTTGTGTCCAGCTTTTTGGTGGAAATGCTTTGAGGCTTCCTTCTGGAGCTAG 90878

Qy    530 rPheArgGlyAspIleAsnValLeuLeuValGlyAspProGlyThrSerLysSerGlnLe 550
Db  90879 TTTCCGAGGCGACATCAATATTTTGCTTGTTGGTGATCCTGGAACAAGTAAATCCCAGCT 90938

Qy    550 uLeuGlnTyrValHisLysIleAlaProArgGlyIleTyrThrSerGlyArgGlySerSe 570
Db  90939 TCTCCAATACATGCACAAACTGTCTCCTCGTGGCATTTATACAAGTGGCAGAGGAAGTTC 90998

Qy    570 rAlaValGlyLeuThrAlaTyrValThrLysAspProGluThrArgGluThr-------- 587
Db  90999 AGCTGTTGGCCTTACTGCATACGTTACCAAGGATCCTGAAACTGGTGAAACTGTAAGTCT 91058

Qy    587 ------------------------------------------------------------ 587
Db  91059 ACATGAATCCTTTATTTCGCAACTTAATTTAACTCGTTTTAATGTAATCTGATTGATGA 91118

Qy    588 ------------------------ValLeuGluSerGlyAlaLeuValLeuSerAspAr 599
Db  91119 CTTCAACCATGACATCTCATTTCAGGTTCTTGAGAGTGGAGCGCTTGTTTTGAGTGACAA 91178

Qy    599 gGlyIleCysCysIleAspGluPheAspLysMetSerAspAsnAlaArgSerMetLeuHi 619
Db  91179 AGGTGTTTGTTGTATTGATGAATTTGATAAGATGTCTGATAATGCTCGAAGCATGTTACA 91238

Qy    619 sGlu-------------------------------------------------------- 620
Db  91239 TGAGGTCTCTTGTAATGCTGCGCCCCAGAATTTGTATCACACAGTACTATTTAGTCAATG 91298

Qy    621 ----------------------------ValMetGluGlnGlnThrValSerValAlaL 631
Db  91299 CTAAACTGTCCTTTTTCAAAATATTGCAGGTGATGGAACAACAGACTGTCTCCATTGCCA 91358

Qy    631 ysGlyGlyIleIleAlaSerLeuAsnAlaArgThrSerValLeuAlaCysAlaAsnProS 651
Db  91359 AGGCTGGAATAATTGCATCTTTAAATGCCAGAACATCAGTTCTAGCATGTGCAAATCCTA 91418

Qy    651 erGlySerArgTyrAsnAlaArgLeuSerValIleAspAsnIleGlnLeuProProThrL 671
Db  91419 CTGAATCACGTTATAATCCAAGGCTCTCTGTGATTGACAATATCCATCTTCCTCCAACAC 91478

Qy    671 euLeuSer---------------------------------------------------- 673
Db  91479 TGCTTTCTAGGTGAGTGCAATGCAATGTGAATGAGCAATTCTTTCAGGACAAACAGATCA 91538

Qy    674 ------------------------------------------------ArgPheAspLeu 677
Db  91539 ATAATAGAAGTGTTTTAACTCTTGATTGTTTTTATTTTTTTCAAATTCAGGTTTGACCTC 91598

Qy    678 IleTyrLeuMetLeuAspLysProAspGluGlnAsnAspArgArgLeuAlaArgHisLeu 697
Db  91599 ATTTATCTGATATTGGACAAGGCAGATGAGCAAACTGATAGACGCCTGGCTAAGCATATT 91658

Qy    698 ValAlaLeuHisTyrGluAsnTyrGluVal-SerLysGln-------------------- 710
Db  91659 GTTTCGTTGCATTTTGAGAATCCAAACGTAAACACAACAACTCCTTTTCTTCCTGAAGTT 91718

Qy    710 ------------------------------------------------------------ 710
Db  91719 TGTACTTTTTTACTTATTTGATATAAACTGCTAACCGTCTTGCAAATCAGAACACTGTTG 91778

Qy    710 ------------------------------------------------------------ 710
Db  91779 TTAGCTTTAGCCTATTTCTTATTAATTGTGCATTCTTTCTGCTATCTAGATAGAGGAGCT 91838

Qy    711 -AspAlaLeuAspLeuGlnThrLeuThrAlaTyrIleThrTyrAlaArgGlnHisValHi 730
Db  91839 CGAGGTCTTGGATTTGCCAACACTTGTAGCCTACATAAGTTATGCAAGGAAGCATATACA 91898

Qy    730 sProThrLeuSerAspGluAlaAlaGluAspLeuIleAsnGlyTyrValGluMetArgGl 750
Db  91899 ACCACAGTTATCTGATGAAGCTGCAGAAGAATTGACTCGCGGCTATGTTGAGATGAGGAA 91958

Qy    750 nLysGlyAsnPheProGlySerSerLysLys----------------------------- 760
Db  91959 AAGAGGAAACAGCCCTGGTAGCAGAAAGAAGGTAGTGTAGGCTTTTAATATGATATTTT 92018

Qy    761 ------------------------------------------------ValIleThrAl 764
Db  92019 TAGCCATTAAATGGTACAAGTGACTGATGCTTTTTTGCTTATTGTATAGGTCATAACTGC 92078

Qy    764 aThrProArgGlnLeuGluSerMetIleArgIleSerGluAlaLeuAlaArgMetArgPh 784
Db  92079 GACAGCTCGACAAATTGAGAGCTTGATTCGGCTCAGTGAAGCACTGGCCCGAATGCGATT 92138

Qy    784 eSerGluVal-------------------------------------------------- 787
Db  92139 CTCTGAAATGGTAATGATCATTCCTTATTCCTAGTTGTCTTGGCTTCACAGCATAATTCT 92198

Qy    788 ------------------------------------------ValGluLysValAspAl 793
Db  92199 GGTTACTTTGGTACTCATGATTTCATTCTTCACTTGTCTACAGGTTGAAGTACAAGATGT 92258

Qy    793 aAlaGluAlaValArgLeuLeuAspValAlaLeuGlnGlnSerAlaThrAspHisAlaTh 813
Db  92259 TGTAGAGGCCTTCAGGCTTCTCGAAGTTGCCATGCAGCAATCGGCAACTGATCATGCCAC 92318

Qy    813 r----------------------------------------------------------- 813
Db  92319 TGGTCAGACAATTCTGCTTAGCGCCATTTCCTATGCATTTTTTTCTTGTAACCATAACAT 92378

Qy    814 -----------------------------------GlyThrIleAspMetAspLeuIleTh 822
Db  92379 AAAGTGCTGACATTTTTATCAATCAACCACTAAAGGTACAATCGATATGGATCTTATCAT 92438

Qy    822 rThrGlyValSerAlaSerGluArgIleArgArgAlaAsnLeuLeuAlaAlaLeuArgGl 842
Db  92439 GACTGGAATATCTGCGAGCGAAAGGCAGAGGCGGGACAATCTCGTTGCAGCAACCCGCAA 92498

Qy    842 uLeuIleAlaAspLysIleSerProGly-------------------------------- 851
Db  92499 CCTTGTGATGGAGAAAATGCAGCTTGGAGGGCCCTCAGTGCGAATGATTGAGGTAAAGAA 92558

Qy    852 -------------------------------------SerSerSerGlyLeuLys----- 857
Db  92559 CCTTCAATTAGTGTCCTTTGTTCTCATCATTGTTTGCCAAACCAGTGGACTAAGTTATTC 92618

Qy    858 ----------ThrSerGlnLeuLeuGluAspIleArgSerGlnSerSerValAspValSe 874
Db  92619 TTCCTGATTGATTTCGCAGTTGCTGGAAGAAATTAGGAAGCAGAGCTCTATGGAAGTTCA 92678

Qy    874 rLeuGlnAspIle----------------------------------------------- 878
Db  92679 TCTGCATGATGTAAGTTGGAGATCTAGTCTGCATTGCTTGCATTGACCTGATACCCTGGG 92738

Qy    879 ------------------------------------------LysAsnAlaLeuGlySerLe 885
Db  92739 AGGCTCTAACCCATTCCTGAATTTTCCATCTTCTCAGCTTCGCGGTGCTCTTGGCACTCT 92798

Qy    885 uGlnGlyGluGlyPheLeuThrValHisGlyAspIleValLysArgVal 901
Db  92799 GATGACAGAGGGTGCCGTAGTCATCCATGGAGACACGCGTCAAGAGGGTC 92847
```

RESULT 5
AP004232
LOCUS AP004232 187471 bp DNA linear PLN 19-JAN-2005
DEFINITION Oryza sativa (japonica cultivar-group) genomic DNA, chromosome 1,
BAC clone:OSJNBa0051H17.
ACCESSION AP004232 BA000010
VERSION AP004232.4 GI:20161778
KEYWORDS .

```
SOURCE      Oryza sativa (japonica cultivar-group)
  ORGANISM  Oryza sativa (japonica cultivar-group)
            Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
            Spermatophyta; Magnoliophyta; Liliopsida; Poales; Poaceae;
            Ehrhartoideae; Oryzeae; Oryza.
REFERENCE   1
  AUTHORS   Sasaki,T., Matsumoto,T., Yamamoto,K., Sakata,K., Baba,T.,
            Katayose,Y., Wu,J., Niimura,Y., Cheng,Z., Nagamura,Y.,
            Antonio,B.A., Kanamori,H., Hosokawa,S., Masukawa,M., Arikawa,K.,
            Chiden,Y., Hayashi,M., Okamoto,M., Ando,T., Aoki,H., Arita,K.,
            Hamada,M., Harada,C., Hijishita,S., Honda,M., Ichikawa,Y.,
            Idonuma,A., Iijima,M., Ikeda,M., Ikeno,M., Ito,S., Ito,T., Ito,Y.,
            Ito,Y., Iwabuchi,A., Kamiya,K., Karasawa,W., Katagiri,S.,
            Kikuta,A., Kobayashi,N., Kono,I., Machita,K., Maehara,T.,
            Mizuno,H., Mizubayashi,T., Mukai,Y., Nagasaki,H., Nakashima,M.,
            Nakama,Y., Nakamichi,Y., Nakamura,M., Namiki,N., Negishi,M.,
            Ohta,I., Ono,N., Saji,S., Sakai,K., Shibata,M., Shimokawa,T.,
            Shomura,A., Song,J., Takazaki,Y., Terasawa,K., Tsuji,K., Waki,K.,
            Yamagata,H., Yamane,H., Yoshiki,S., Yoshihara,R., Yukawa,K.,
            Zhong,H., Iwama,H., Endo,T., Ito,H., Hahn,J.H., Kim,H.I., Eun,M.Y.,
            Yano,M., Jiang,J. and Gojobori,T.
  TITLE     The genome sequence and structure of rice chromosome 1
  JOURNAL   Nature 420 (6913), 312-316 (2002)
   PUBMED   12447438
REFERENCE   2  (bases 1 to 187471)
  AUTHORS   Sasaki,T., Matsumoto,T. and Yamamoto,K.
  TITLE     Direct Submission
  JOURNAL   Submitted (03-OCT-2001) Takuji Sasaki, National Institute of
            Agrobiological Sciences, Rice Genome Research Program; Kannondai
            2-1-2, Tsukuba, Ibaraki, 305-8602, Japan
            (E-mail:tsasaki@nias.affrc.go.jp, URL:http://rgp.dna.affrc.go.jp/,
            Tel:81-298-38-7441, Fax:81-298-38-7468)
COMMENT     On Apr 16, 2002 this sequence version replaced gi:19110523.
            Genes were predicted from the integrated results of the following:
            GENSCAN (http://CCR-081.mit.edu/GENSCAN.html), FGENESH
            (http://www.softberry.com/), GeneMark.hmm
            (http://opal.biology.gatech.edu/GeneMark/), GlimmerM
            (http://www.tigr.org/tdb/glimmerm/glmr_form.html), RiceHMM
            (http://rgp.dna.affrc.go.jp/RiceHMM/), SplicePredictor
            (http://bioinformatics.iastate.edu/cgi-bin/sp.cgi), sim4
            (http://globin.cse.psu.edu/html/docs/sim4.html), gap2
            (http://www.tigr.org/software/glimmerm/), BLASTN and BLASTX. The
            genomic sequence was searched against NCBI NonRedundant Protein
            database, nr (ftp://ncbi.nlm.nih.gov/blast/db) and the cDNA
            sequence database at RGP or DDBJ. Protein homologies of the coding
            regions were searched against NCBI NonRedundant Protein database
            with BLASTP. ESTs represent the identified cDNA sequences using
            BLASTN with the corresponding DDBJ accession no. and RGP clone ID.
            Full-length cDNAs represent the identified cDNA sequences using
            BLASTN with the corresponding DDBJ accession no.
            A gene with identity or significant homology to a protein is
            classified based on the protein name to indicate the homology level
            such as same name, 'putative-' and '-like protein'.  A gene without
            significant homology to any protein but with full-length cDNA or
            EST homology (covering almost the entire length of partial
            sequence) is classified as an 'unknown' protein. A gene predicted
            by two or more gene prediction programs is classified as a
            'hypothetical' protein according to IRGSP standard. A gene
            predicted by a single gene prediction program is also classified as
            a probable 'hypothetical' protein and is included as a
            miscellaneous feature of the sequence.
            The orientation of the sequence is from -21M13 to M13rev of the BAC
            clone. This sequence of OSJNBa0051H17 clone has an overlap with
            OSJNBa0065J17 (DDBJ: AP004233) clone at 5' end and with
            OSJNBa0026J14 (DDBJ: AP004231) clone at 3' end. Detailed
            information on overlap and assembly quality together with
            annotation of this entry is available at
            http://rgp.dna.affrc.go.jp/GenomeSeq.html.
FEATURES             Location/Qualifiers
     source          1. .187471
                     /organism="Oryza sativa (japonica cultivar-group)"
                     /mol_type="genomic DNA"
                     /cultivar="Nipponbare"
                     /db_xref="taxon:39947"
                     /chromosome="1"
                     /clone="OSJNBa0051H17"
     gene            3101. .4772
                     /gene="OSJNBa0051H17.1"
     mRNA            join(<3101. .3168,3216. .3273,3487. .3700,4630. .>4772)
                     /gene="OSJNBa0051H17.1"
                     /note="start and end point are not identified"
     CDS             join(3101. .3168,3216. .3273,3487. .3700,4630. .4772)
                     /gene="OSJNBa0051H17.1"
                     /note="predicted by GENSCAN etc."
                     /codon_start=1
                     /product="hypothetical protein"
                     /protein_id="BAB90694.1"
                     /db_xref="GI:20161779"
                     /translation="MWRLSAAAGDVVLKQIEQASRPGVGGGPCRTPPVRQASRRRVVV
                     IRSREQTVALINPTYFYFADTHGERIDYNTQQKGIRDEAAAGALRRHHHRALLRRRRL
                     GRRPRRPLRPQGRYMEKMTLQNYRNVPAGTNAQACLAVAKELNYPVIDIWTKMQQFPD
                     "
     gene            complement(join(5163. .5279,7637. .7845,7935. .8050,
                     8149. .8210))
                     /gene="OSJNBa0051H17.2"
     misc_feature    complement(join(5163. .5279,7637. .7845,7935. .8050,
                     8149. .8210))
                     /gene="OSJNBa0051H17.2"
                     /note="transposase-like protein"
     gene            join(9375. .9530,10947. .11081)
                     /gene="OSJNBa0051H17.3"
     misc_feature    join(9375. .9530,10947. .11081)
                     /gene="OSJNBa0051H17.3"
                     /note="hypothetical ORF
                     predicted by GENSCAN
                     this category is not included in IRGSP standard"
     gene            11904. .12794
                     /gene="OSJNBa0051H17.4"
     misc_feature    11904. .12794
                     /gene="OSJNBa0051H17.4"
                     /note="probably inactive due to 5' exon missing in CDS
                     pseudogene, TNP2-like transposon protein"
                     /pseudo
     gene            13002. .15220
                     /gene="OSJNBa0051H17.5"
     misc_feature    13002. .15220
                     /gene="OSJNBa0051H17.5"
                     /note="probably inactive due to including stop codon(s) in
                     CDS
                     pseudogene, Rim2 protein"
                     /pseudo
     gene            15984. .16427
                     /gene="OSJNBa0051H17.6"
     mRNA            <15984. .>16427
                     /gene="OSJNBa0051H17.6"
                     /note="start and end point are not identified"
     CDS             15984. .16427
                     /gene="OSJNBa0051H17.6"
                     /note="predicted by GlimmerM etc."
                     /codon_start=1
                     /product="hypothetical protein"
                     /protein_id="BAD88089.1"
                     /db_xref="GI:57900494"
                     /translation="MRPRVLHQHLEVFPPISASPSPPLPRAFPSPPSSLAVVSRSSLH
                     PPPPSPTPPPHLWGERRRCLSFSLSRSLLSLFLQHPPRRLVTDAATDIDAADPDSASR
                     RCRPRLRRGRHRRRRRHPHAAAPASSSPRRHATPLPPLPPRHCRT"
     gene            17301. .20737
                     /gene="OSJNBa0051H17.7"
     misc_feature    17301. .20737
                     /gene="OSJNBa0051H17.7"
                     /note="probably inactive due to including stop codon(s) in
                     CDS
                     pseudogene, putative hydroxyproline-rich glycoprotein"
                     /pseudo
     gene            complement(join(21536. .21593,21725. .21819))
                     /gene="OSJNBa0051H17.8"
```

```
     misc_feature    complement(join(21536. .21593,21725. .21819))
                     /gene="OSJNBa0051H17.8"
                     /note="hypothetical ORF
                     predicted by RiceHMM
                     this category is not included in IRGSP standard"
     gene            26939. .29186
                     /gene="OSJNBa0051H17.9"
     misc_feature    26939. .29186
                     /gene="OSJNBa0051H17.9"
                     /note="probably inactive due to including stop codon(s) in
                     CDS
                     pseudogene, P450"
                     /pseudo
     gene            complement(33543. .33875)
                     /gene="OSJNBa0051H17.10"
     misc_feature    complement(33543. .33875)
                     /gene="OSJNBa0051H17.10"
                     /note="hypothetical ORF
                     predicted by GlimmerM
                     this category is not included in IRGSP standard"
     gene            35632. .36027
                     /gene="OSJNBa0051H17.11"
     mRNA            join(<35632. .35645,35814. .>36027)
                     /gene="OSJNBa0051H17.11"
                     /note="start and end point are not identified"
     CDS             join(35632. .35645,35814. .36027)
                     /gene="OSJNBa0051H17.11"
                     /note="predicted by GENSCAN etc."
                     /codon_start=1
                     /product="hypothetical protein"
                     /protein_id="BAB90700.1"
                     /db_xref="GI:20161785"
                     /translation="MGGRDAGGQRAGWPTACRRGGPGGLEDSQAARRPGRRGGAVLGG
                     GAAGEVPGGLVAGFICDEFVIRICVFDLCVM"
     gene            complement(41636. .42103)
                     /gene="OSJNBa0051H17.12"
     misc_feature    complement(41636. .42103)
                     /gene="OSJNBa0051H17.12"
                     /note="retroelement-like protein"
     gene            complement(42773. .43616)
                     /gene="OSJNBa0051H17.13"
     mRNA            complement(join(<42773. .42832,43509. .>43616))
                     /gene="OSJNBa0051H17.13"
                     /note="start and end point are not identified"
     CDS             complement(join(42773. .42832,43509. .43616))
                     /gene="OSJNBa0051H17.13"
                     /note="predicted by GeneMark.hmm etc."
                     /codon_start=1

Alignment Scores:
Pred. No.:              8.92e-62        Length:         187471
Score:                  1937.50         Matches:        530
Percent Similarity:     41.94%          Conservative:   133
Best Local Similarity:  33.52%          Mismatches:     173
Query Match:            42.04%          Indels:         751
DB:                     15              Gaps:           31

US-10-768-511-8 (1-901) x AP004232 (1-187471)

Qy         11 AlaValSerSerProTyrProSerGlnSerGluGlyValSerThrProLeuProGlnVal 30
              |||||||||||||||                         |||:::|||||||||||
Db      94668 GCCGTGTCGTCGCCG------------GATGTGCGACCGTCGAGCCCTCTCCCGGCGACC 94715

Qy         31 ThrSerProSerPheAspAsnAlaAlaSerProValAlaGlyArgArgAlaValArgGln 50
                                      |||::::::  |||   :::||||||||||
Db      94716 -----------------AACTCCTCCCCGCCTCAGTCGGGGCGGCGCGGC--------- 94748

Qy         51 ThrProThrSerAlaValArgArgArgGlyArgGluThrAspSerAlaArgArgArgArg 70
                                                          ||||||||||||
Db      94749 -------------------------------------------GGGGGGCGGCGCCGGCGT 94766

Qy         71 SerArgSerArgSerLeuGlyAsnSerValTyrSerSerProTyrAspAlaGlyThrPro 90
                                                 :::|||||||||   :::   :::|||
Db      94767 GGTTCC-------------------------GCCAGCCCGTACCCGTCG---TCCCCG 94796

Qy         91 GlyThrProGlyThrProValAlaThrProValTyrAlaThrProValGlyThrProMet 110
                      |||                         :::   ||||||
Db      94797 TCCCTCGGCGGG--------------------TTCGAGACGCCGCCGCACCCGGGCCGC 94835

Qy        111 GlyThrProSerPheHisArgGlyThrProGlnTyrLysGlnArgSerGluLeuGlySer 130
                 |||||||||          |||         :::||||||    ||||||
Db      94836 CGCACGCCCTCC------GGCGGCGCGGCGGCGCGGCAGCAGCGGCA-GAACTGGACGGG 94888

Qy        131 GlnGlyLysPro--------------------LeuHisArgArgArgArgSerGlnSer 143
                       |||                      :::|||||||||||||
Db      94889 GGGGCGGTTCCCGCCGACCCCGTCCACCCCCATGTCCACCGACGACGTCC-CGCTCTCCT 94947

Qy        144 ArgGluProGlyHisArgSerProSerArgGluProSerAla--AspGlyArgProSerG 163
              |||    ||||||    |||:::|||    |||        :::|||     |||
Db      94948 CGGAGGCCGGGGACGAGGACACCCCCGAGACCGACGGCGGCGGCGGCGGCGGCGCGGGCG 95007

Qy        163 luSerAlaGluProAspAspThrLeuGlyGlyGluTyrAlaTyrValTrpGlyThrAsnV 183
                 |||   |||                              :::||||||||||||||:
Db      95008 CCGACGCCACGCCG-----------------------GTGTTCGTCTGGGGCACCAACA 95043

Qy        183 alAsnIleProAspValLeuArgAlaIleArgArgPheLeuHisAsnTyrArgSerSerA 203
              ::::::::    ||||||        ||||||   |||||||||    ::::::|||
Db      95044 TCAGCGTGCAGGACGTCAACGCCGCCATCCTCCGGTTCCTCCGCCACTTCCGC---GACC 95100

Qy        203 laHisAspLeu----------------------AsnSerLysTyrIleGlnIleIleG 215
                 |||                                 ||||||::::::    |||
Db      95101 CGCGCGACGCCGGCCGCGTCGACCCGGTCATGGACGAGGGCAAGTACATGCGCGCCATCC 95160

Qy        215 luGluThrValGluArgGlu---GluAspThrLeuAsnIleAspMetSerAspIleTyrA 234
                 :::|||    |||        ::::::|||:::::::::        |||::::::|
Db      95161 ACCGCATCCTCGAGCTCGAGGGCGGGGAGTCGCTCGACGTCAACGCCCACGACGTGTTCG 95220

Qy        234 spHisAspProAspLeuTyrAlaLysIleValArgTyrProLeuAspIleIleProLeuL 254
              ||||||||||||||||||    |||:::||||||||||||||||:::::::::    :::
Db      95221 ACCACGACCCGGACCTCTACGGCAAGATGGTCAGGTACCCGCTCGAGGTGCTCGCCATCT 95280

Qy        254 euAspThrGluCysGlnGluValAlaThrSerLeuLeuProThrPheGluLysHisIleG 274
                 |||          ::::::         :::   |||    |||||||||||||||:
Db      95281 TTGACATCGTGCTCATGGACCTCGTCGCGCGCATCGAGCCCCTCTTCGAGAAGCACATCC 95340

Qy        274 luAlaArgProPheAsnLeuLysAlaSerValHisMetArgGluLeuAsnProSer---- 292
              ::    |||    :::|||||||||:::||||||    :::|||    |||||||||
Db      95341 AGACCAGGATCTACAACCTCAAGTCCTCGGTTTGCTTGAGGAATCTCAACCCTTC-TGGT 95399

Qy        292 ------------------------------------------------------------ 292

Db      95400 GATTGCGCACTGCTCTGCTCCTTTAGCTACTTGTGGTCGGCAATTCTTTTATATGATTTC 95459

Qy        293 --------------------AspIleAspLysLeuValSerValLysGlyMetValIleA 306
                                  ||||||:::|||:::||||||:::|||||||||:::||||
Db      95460 TGATTTTGCGTTTCGTTGCAGATATCGAGAAGATGGTGTCCATCAAGGGTATGATAATTC 95519

Qy        306 rgCysSerSerIleIleProGluIleLysGlyAlaPhePheLysCysLeuValCysGly- 325
              ||||||||||||:::|||||||||:::|||    |||    |||:::|||||||||||||
Db      95520 GGTGCAGCTCGGTGATTCCAGAGCTCAAGGAGGCTGTGTTCCGCTGCTTGGTTTGCGGGT 95579

Qy        326 --HisSerProProLeuVal---------------------------------------- 331
                :::|||    |||::::::
Db      95580 TCTACTCTGAGCCTGTAATGGTTGATCGAGGTATATATATATCAGAACTAATGTGGATAC 95639

Qy        331 ------------------------------------------------------------ 331

Db      95640 CTTTTTCACACTGCGTATTCCATATGCTCTTGTTTTAAGTGAACTTCAGGTGATCAATTA 95699

Qy        331 ------------------------------------------------------------ 331

Db      95700 ATCAGATTGCGAGCCTTAATTTTGCTCATAGAGAGTTACATATTGACAACTCATTGTTAA 95759

Qy        331 ------------------------------------------------------------ 331

Db      95760 CTTCATTTGTACAAATCATTCTTATCATGAAAGATGATAAATTCAAGCTTCATTTGACAT 95819
```

```
Qy        331 ------------------------------------------------------------ 331
Db      95820 GTTATATGTCATTACTATTATCATTTATGCATTTTCTTTTATTTTGTTTGCAAAAACTTT 95879
Qy        331 ------------------------------------------------------------ 331
Db      95880 TCATTTTGTTGGGATACACAACATTTAGGAATTTTATTTATCATACCACATTGCTGAGGA 95939
Qy        331 ------------------------------------------------------------ 331
Db      95940 ACTGATAAATAGTGATACTAATTGAGTACGCAAATCAAACATCCATACTGCTATTGTCCA 95999
Qy        331 ------------------------------------------------------------ 331
Db      96000 ATCCAAGACTCCGACTTAGTCTGTCTGTCATAAAGTAACACTTTGCACTGTGGTAGTAAT 96059
Qy        331 ------------------------------------------------------------ 331
Db      96060 TGTAAGTAAAAACATGCTTATGTTCTTAACTTACATTATATTCTGGTGGTTAATGTTTGA 96119
Qy        331 ------------------------------------------------------------ 331
Db      96120 AGAGATTGGTCATGCAGATCCTATGAAAACACATTAGTGGCCTAGTGGGAGTAATGACTA 96179
Qy        332 -----------------------------------------ThrVal-ValLysGlyArg 337
Db      96180 AAACTCCATCAGTTTCTGTTGTTTTATCATGTAACTAATTCACTCTATTTCTTAGGGAGG 96239
Qy        338 ValGluGluProThrArgCysGluLysProGluCysAlaAlaArgAsnAlaMetSerLeu 357
Db      96240 GTTACTGAGCCGCACATCTGTCAGAAAGAACAATGTAAAGCTACAAACTCTATGACTCTT 96299
Qy        358 IleHisAsnArgCysThr------------------------------------------ 363
Db      96300 GTGCATAACCGATGCAG-GTAGCTATATCCTACTTGTAACCCCTTTTTCTTTATAGGATT 96358
Qy        364 ------------------------------------------------------PheAla 365
Db      96359 TTTGACAGATATGTAGTCCTGACCTTTGGATGGAACATTTTATTTGTAACTAGGTTTGCA 96418
Qy        366 AsnLysGlnIleValArgLeuGlnGluThrProAspAlaIleProGluGlyGluThrPro 385
Db      96419 GATAAGCAGATCATAAAGTTGCAGGAAACACCAGATGAGATACCAGAAGGTGGCACTCCA 96478
Qy        386 HisThrValSerMetCysLeuTyrAsnThrMetValAspAlaValLysProGlyAspArg 405
Db      96479 CATACTGTTAGTGTCTTGATGCACGATAAGCTTGTAGATGCTGGAAAGCCTGGAGACAGG 96538
Qy        406 IleGluVal--------------------------------------------------- 408
Db      96539 GTTGAGGTTAGCAGTGAATCTTTGTTTGGTCATTAATACAAGAAAATAATTTTGTTTGTG 96598
Qy        409 ---------------------------------------------------ThrGlyVal 411
Db      96599 GTTCTTATTCACATCGTAAAAATCTTGATTTTTTTCTTTGATTTATAGATAACTGGGATA 96658
Qy        412 PheLysAlaMetAlaValArgValGlyProAsnGlnArgThrLeuArg------------ 427
Db      96659 TACAGAGCCATGAGTATTAGAGTTGGCCCAACTCAGAGGACAGTGAAGTCGATTTTCAAG 96718
Qy        427 ------------------------------------------------------------ 427
Db      96719 GTAGAGGATGTGCAATGTTATCCTACTTGCTATCTTAAATTCCTTTTCTGCACAGGAAAA 96778
Qy        428 ---------------------------------------AlaLeuTyrLysThrTyrIle 434
Db      96779 AAGTTTTAAATTATCTACTCAAGCATCGTCTAACAGTTTCTTTAAT-CAGACGTACATT 96837
Qy        435 AspCysValHisValLysLysSerAspArgGlyArgLeuGlnThrGluAspProMetGlu 454
Db      96838 GATTGCCTTCACATAAAGAAGACAGACAAGTCTAGACTTCATGTTGAGGACTCCATGGAA 96897

Qy        455 MetAspLysGluAsn---------------------------------AspMetTyrAla 463
Db      96898 ACTGATAACCCCAATGCTAACAAGACAACTGAAGATGATTTTCTCAGAGATAAGGTATCC 96957
Qy        464 GlyTyrHisGluSerAspThrSerGluAla---------------------------Ala 474
Db      96958 CTCAATCACCTTACCTCTCTGGGATA-GCACTAAGTGTAAATGTGTTTAACACTGTCCTT 97016
Qy        475 AsnGluAlaLysIleGlnLysLeuLysGluLeuSerLysLeuProGlyIleTyrAspArg 494
Db      97017 TCTGTATTGCAGGTTGAGAAATTAAAAGAGTTGTCAAAGTTGCCAGATATATATGACAGA 97076
Qy        495 LeuSerArgSerLeuAlaProSerIleTrpGluLeuGluAspIleLysLysGly------ 512
Db      97077 TTAACTAGGTCATTGGCTCCAAACATATGGGAGCTGGACGACGTTAAAAGGGGCCTCCTT 97136
Qy        512 ------------------------------------------------------------ 512
Db      97137 TGCCAGGTAAGATCATCTATCTGTTTGAATGTTTTTGCACTAATAGCCTAGTATAACAAG 97196
Qy        513 ------LeuLeuCys-GlnLeuPheGlyGlyLysAlaLysLysIleProSerGlyAlaSe 530
Db      97197 CTCTTACTGTTGTGTCCAGCTTTTTGGTGGAAATGCTTTGAGGCTTCCTTCTGGAGCTAG 97256
Qy        530 rPheArgGlyAspIleAsnValLeuLeuValGlyAspProGlyThrSerLysSerGlnLe 550
Db      97257 TTTCCGAGGCGACATCAATATTTTGCTTGTTGGTGATCCTGGAACAAGTAAATCCCAGCT 97316
Qy        550 uLeuGlnTyrValHisLysIleAlaProArgGlyIleTyrThrSerGlyArgGlySerSe 570
Db      97317 TCTCCAATACATGCACAAACTGTCTCCTCGTGGCATTTATACAAGTGGCAGAGGAAGTTC 97376
Qy        570 rAlaValGlyLeuThrAlaTyrValThrLysAspProGluThrArgGluThr-------- 587
Db      97377 AGCTGTTGGCCTTACTGCATACGTTACCAAGGATCCTGAAACTGGTGAAACTGTAAGTCT 97436
Qy        587 ------------------------------------------------------------ 587
Db      97437 ACATGAATCCTTTATTTCGCAACTTAATTTAACTCGTTTTAATGTAATCTGATTGATGA 97496
Qy        588 -------------------------ValLeuGluSerGlyAlaLeuValLeuSerAspAr 599
Db      97497 CTTCAACCATGACATCTCATTTCAGGTTCTTGAGAGTGGAGCGCTTGTTTTGAGTGACAA 97556
Qy        599 gGlyIleCysCysIleAspGluPheAspLysMetSerAspAsnAlaArgSerMetLeuHi 619
Db      97557 AGGTGTTTGTTGTATTGATGAATTTGATAAGATGTCTGATAATGCTCGAAGCATGTTACA 97616
Qy        619 sGlu-------------------------------------------------------- 620
Db      97617 TGAGGTCTCTTGTAATGCTGCGCCCCAGAATTTGTATCACACAGTACTATTTAGTCAATG 97676
Qy        621 -----------------------------ValMetGluGlnGlnThrValSerValAlaL 631
Db      97677 CTAAACTGTCCTTTTTCAAAATATTGCAGGTGATGGAACAACAGACTGTCTCCATTGCCA 97736
Qy        631 ysGlyGlyIleIleAlaSerLeuAsnAlaArgThrSerValLeuAlaCysAlaAsnProS 651
Db      97737 AGGCTGGAATAATTGCATCTTTAAATGCCAGAACATCAGTTCTAGCATGTGCAAATCCTA 97796
Qy        651 erGlySerArgTyrAsnAlaArgLeuSerValIleAspAsnIleGlnLeuProProThrL 671
Db      97797 CTGAATCACGTTATAATCCAAGGCTCTCTGTGATTGACAATATCCATCTTCCTCCAACAC 97856
Qy        671 euLeuSer---------------------------------------------------- 673
Db      97857 TGCTTTCTAGGTGAGTGCAATGCAATGTGAATGAGCAATTCTTTCAGGACAAACAGATCA 97916
Qy        674 ------------------------------------------------ArgPheAspLeu 677
Db      97917 ATAATAGAAGTGTTTTAACTCTTGATTGTTTTTATTTTTTTCAAATTCAGGTTTGACCTC 97976
Qy        678 IleTyrLeuMetLeuAspLysProAspGluGlnAsnAspArgArgLeuAlaArgHisLeu 697
```

```
                 |||||||||:::|||||||||  |||||||||  |||||||||||||||:::|||:::
Db       97977 ATTTATCTGATATTGGACAAGGCAGATGAGCAAACTGATAGACGCCTGGCTAAGCATATT 98036

Qy         698 ValAlaLeuHisTyrGluAsnTyrGluVal-SerLysGln-------------------- 710
               |||:::||||||:::||||||     |||  :::::::|||
Db       98037 GTTTCGTTGCATTTTGAGAATCCAAACGTAAACACAACAACTCCTTTTCTTCCTGAAGTT 98096

Qy         710 ------------------------------------------------------------ 710

Db       98097 TGTACTTTTTTACTTATTTGATATAAACTGCTAACCGTCTTGCAAATCAGAACACTGTTG 98156

Qy         710 ------------------------------------------------------------ 710

Db       98157 TTAGCTTTAGCCTATTTCTTATTAATTGTGCATTCTTTCTGCTATCTAGATAGAGGAGCT 98216

Qy         711 -AspAlaLeuAspLeuGlnThrLeuThrAlaTyrIleThrTyrAlaArgGlnHisValHi 730
                 :::   |||||||||  ||||||  |||||||||:::|||||||||:::|||:::
Db       98217 CGAGGTCTTGGATTTGCCAACACTTGTAGCCTACATAAGTTATGCAAGGAAGCATATACA 98276

Qy         730 sProThrLeuSerAspGluAlaAlaGluAspLeuIleAsnGlyTyrValGluMetArgGl 750
                 |||   ||||||||||||||||||||||:::|||     ||||||||||||||||||::
Db       98277 ACCACAGTTATCTGATGAAGCTGCAGAAGAATTGACTCGCGGCTATGTTGAGATGAGGAA 98336

Qy         750 nLysGlyAsnPheProGlySerSerLysLys----------------------------- 760
               ::::||||||   |||||||||   ||||||
Db       98337 AAGAGGAAACAGCCCTGGTAGCAGAAAGAAGGTAGTGTAGGCTTTTTAATATGATATTTT 98396

Qy         761 ---------------------------------------------------ValIleThrAl 764
                                                                 ||||||||||
Db       98397 TAGCCATTAAATGGTACAAGTGACTGATGCTTTTTTGCTTATTGTATAGGTCATAACTGC 98456

Qy         764 aThrProArgGlnLeuGluSerMetIleArgIleSerGluAlaLeuAlaArgMetArgPh 784
               ||||   ||||||:::|||||||:::||||||:::||||||||||||||||||||||||
Db       98457 GACAGCTCGACAAATTGAGAGCTTGATTCGGCTCAGTGAAGCACTGGCCCGAATGCGATT 98516

Qy         784 eSerGluVal-------------------------------------------------- 787
               |||||||:::
Db       98517 CTCTGAAATGGTAATGATCATTCCTTATTCCTAGTTGTCTTGGCTTCACAGCATAATTCT 98576

Qy         788 --------------------------------------------ValGluLysValAspAl 793
                                                          ||||||    |||
Db       98577 GGTTACTTTGGTACTCATGATTTCATTCTTCACTTGTCTACAGGTTGAAGTACAAGATGT 98636

Qy         793 aAlaGluAlaValArgLeuLeuAspValAlaLeuGlnGlnSerAlaThrAspHisAlaTh 813
                    ||||||  ||||||||:::||||||:::||||||||||||||||||||||||||
Db       98637 TGTAGAGGCCTTCAGGCTTCTCGAAGTTGCCATGCAGCAATCGGCAACTGATCATGCCAC 98696

Qy         813 r----------------------------------------------------------- 813
               |
Db       98697 TGGTCAGACAATTCTGCTTAGCGCCATTTCCTATGCATTTTTTTCTTGTAACCATAACAT 98756

Qy         814 -----------------------------------GlyThrIleAspMetAspLeuIleTh 822
                                                 |||||||||||||||||||||||||
Db       98757 AAAGTGCTGACATTTTTATCAATCAACCACTAAAGGTACAATCGATATGGATCTTATCAT 98816

Qy         822 rThrGlyValSerAlaSerGluArgIleArgArgAlaAsnLeuLeuAlaAlaLeuArgGl 842
                |||||:::|||||||||||||||   ||||||   ||||||:::||||||   |||
Db       98817 GACTGGAATATCTGCGAGCGAAAGGCAGAGGCGGGACAATCTCGTTGCAGCAACCCGCAA 98876

Qy         842 uLeuIleAlaAspLysIleSerProGly------------------------------ 851
                |||:::   :::|||:::      |||
Db       98877 CCTTGTGATGGAGAAAATGCAGCTTGGAGGGCCCTCAGTGCGAATGATTGAGGTAAAGAA 98936

Qy         852 ---------------------------------------SerSerSerGlyLeuLys----- 857
                                                     :::|||||||||
Db       98937 CCTTCAATTAGTGTCCTTTGTTCTCATCATTGTTTGCCAAACCAGTGGACTAAGTTATTC 98996

Qy         858 ----------ThrSerGlnLeuLeuGluAspIleArgSerGlnSerSerValAspValSe 874
                         |||||||||||||||:::||||||   |||||||||::::::|||
Db       98997 TTCCTGATTGATTTCGCAGTTGCTGGAAGAAATTAGGAAGCAGAGCTCTATGGAAGTTCA 99056

Qy         874 rLeuGlnAspIle--------------------------------------------- 878
                |||   |||:::
Db       99057 TCTGCATGATGTAAGTTGGAGATCTAGTCTGCATTGCTTGCATTGACCTGATACCCTGGG 99116

Qy         879 ---------------------------------------LysAsnAlaLeuGlySerLe 885
                                                     :::   |||||||||:::||
Db       99117 AGGCTCTAACCCATTCCTGAATTTTCCATCTTCTCAGCTTCGCGGTGCTCTTGGCACTCT 99176

Qy         885 uGlnGlyGluGlyPheLeuThrValHisGlyAspIleValLysArgVal 901
                |     ||||||   :::   :::|||||||||   ||||||||||||
Db       99177 GATGACAGAGGGTGCCGTAGTCATCCATGGAGACAGCGTCAAGAGGGTC 99225

RESULT 6
AC116551/c
LOCUS       AC116551               153751 bp    DNA     linear   INV 21-FEB-2004
DEFINITION  Dictyostelium discoideum chromosome 2 map complement(581327-427576)
            strain AX4, complete sequence.
ACCESSION   AC116551 AC116101 AC116549 AC116965
VERSION     AC116551.2  GI:28829639
KEYWORDS    HTG.
SOURCE      Dictyostelium discoideum
  ORGANISM  Dictyostelium discoideum
            Eukaryota; Mycetozoa; Dictyosteliida; Dictyostelium.
REFERENCE   1  (bases 1 to 153751)
  AUTHORS   Gloeckner,G., Eichinger,L., Szafranski,K., Pachebat,J., Dear,P.,
            Lehmann,R., Baumgart,C., Parra,G., April,J.F., Guigo,R., Kumpf,K.,
            Tunggal,B., Cox,E., Quail,M.A., Platzer,M., Rosenthal,A. and
            Noegel,A.A.
  TITLE     Sequence and analysis of chromosome 2 of Dictyostelium discoideum
  JOURNAL   Nature 418 (6893), 79-85 (2002)
   PUBMED   12097910
  REMARK    The Dictyostelium Genome Sequencing Consortium
REFERENCE   2  (bases 1 to 153751)
  AUTHORS   Baumgart,C.
  TITLE     Direct Submission
  JOURNAL   Submitted (29-MAR-2002) Genome Analysis, Institute of Molecular
            Biotechnology, Beutenberstr. 11, Jena 07745, Germany
REFERENCE   3  (bases 1 to 153751)
  AUTHORS   Baumgart,C.
  TITLE     Direct Submission
  JOURNAL   Submitted (04-MAR-2003) Genome Analysis, Institute of Molecular
            Biotechnology, Beutenberstr. 11, Jena 07745, Germany
REFERENCE   4  (bases 1 to 153751)
  AUTHORS   Baumgart,C.
  TITLE     Direct Submission
  JOURNAL   Submitted (12-MAR-2003) Genome Analysis, Institute of Molecular
            Biotechnology, Beutenberstr. 11, Jena 07745, Germany
REFERENCE   5  (bases 1 to 153751)
  AUTHORS   Baumgart,C.
  TITLE     Direct Submission
  JOURNAL   Submitted (21-FEB-2004) Genome Analysis, Institute of Molecular
            Biotechnology, Beutenberstr. 11, Jena 07745, Germany
COMMENT     On or before Feb 21, 2004 this sequence version replaced
            gi:28850309, gi:19807752, gi:19920064, gi:19807784.
            CDS predictions from GeneID do not necessarily reflect true genes.
            Further Information is available from IMB Jena, Department of
            Genome Analysis
            (http://genome.imb-jena.de/dictyostelium/)
            and the Univerity Colonge, Institute for Biochemistry I
            (http://www.uni-koeln.de/dictyostelium/project.shtml
            Funding
            Agency : Deutsche Forschungsgemeinschaft (DFG).
FEATURES             Location/Qualifiers
     source          1..153751
                     /organism="Dictyostelium discoideum"
                     /mol_type="genomic DNA"
                     /strain="AX4"
                     /db_xref="taxon:44689"
                     /chromosome="2"
                     /map="complement(581327-427576)"
     CDS             complement(join(<351..2554,2687..2807))
                     /note="GeneID exon scores (in order of location ranges):
                     70.76, 0.70 - GSCJ_ID dd_02254"
                     /codon_start=1
                     /product="hypothetical protein"
```

```
                     /protein_id="AAO52218.1"
                     /db_xref="GI:28829702"
                     /translation="MNESISDHNEKLFWKVFRNKYLNIEIFKQLESTEIEYDSFTIAS
                     LIKLSVNNSNNNNDNNNKLIEIFKLLMEKEKYLMEGLDLVTISISSRNFEILKILLNK
                     PYSFLVKETDFNDALQKYGDSEIEKSILQLLGNQININTFLIENNFQIKNSIPFFKNT
                     IQFILNNQNNDDNNDDNKNNNNKLIKLLELKIYSIKKDEILNLELIEMQNKLLDLNYF
                     SKENFQNSNIYWFKELYEFKGVKTIFIHLILLVFKIYYKRKSLNIPNEIEIEINEIMN
                     NSLLIKNEEKEKKLLKLLINETQSNLLFKLYYYRYYEIIEGCKLSKIVYRNLKTFNIF
                     KLIESNNNSNNNKEEIIRIIKNFINVYEFKNNENDKQLINKFRNVILNNLNSKEYYHK
                     LIETIGNSQIKTFGLPIEWNINLLDFLLDDEINNKIIIIKDKQVIDWLSNMPRLNNRI
                     FTFDIKTIYYYKRKKNLIQFQSLELIEYAIKKLETNNNNNSNNDNLPPIKLETINWSM
                     VYNNPLKSIIDEFKFQNESIINEDELNDKVTFNLFKFLKFEKNEFTILVQLIKYFIEK
                     VLNKYGQDKSYEILKNFVKIDNLCKFQKETLELLVLSTINLDLIMVMIILNCGIDFTN
                     KQDEILINFWNVHILTNTSIFYIPTINGNINFFKIITHIKPQYIINNIKLWCEQTSLL
                     ECKHLAINHLFKKLIQTKGTTLNQILYIKDFINTINNNNNNNNNNNNNNNNNIIKLDHS
                     KFHSIHYLNLKSPKLILYLIESNQIRDWNYLLIPQ"
     CDS             3515. .5095
                     /note="GeneID exon scores (in order of location ranges):
                     144.31 - GSCJ_ID dd_02253"
                     /codon_start=1
                     /product="similar to Arabidopsis thaliana (Mouse-ear
                     cress). Hypothetical 59.5 kDa protein"
                     /protein_id="AAO52156.1"
                     /db_xref="GI:28829640"
                     /translation="MNNQVKPTLSNKAEIINKWLIDNGVRINNKLIKIVYLGKENNFE
                     QTENTTATTTTSERINDSIVSGLGVISLKELKVDDIVAKIPKSIILSIHTSSISNILE
                     KYKIENNIGTSIALIHEASLGEKSKWYGYISSLPRKVDVPILWDSESRKLLKGTAIED
                     VLNDDDILINQVYADVIESILSKNHPEIFGDKELYSIENFKIANSIISSRAFCVDSYH
                     GDSLVPLADIFNHQTAREHVHIESNGDVCNKCGSIKTCKHRKVVTQHHTVNSTKGKRT
                     HKVAGIPSSKKHIHKGNCCSTTTTKTNEEDKDTIIEEDDEHLYIKVVKGVEANKEVYN
                     TYGDHDNAILLSKYGFLEMDNPCDRLSIDKQLVDKELELIGAENQITRMDLFNRISFY
                     AEMFDIDSRNNHAIENDGRLDDALVCSVGIALAPQSIFQSWKSMSEHKREKYFEQLEA
                     EDIVKQSQLVRNTLLSILNKLFSNYNNSNETTATTTIEQDQEKLSTLKNQREIIAQSL
                     KICEKTLINKSINYYKNL"
     CDS             complement(5235. .7895)
                     /note="GeneID exon scores (in order of location ranges):
                     270.65 - GSCJ_ID dd_03185"
                     /codon_start=1
                     /product="similar to essential for initiation of DNA
                     replication; homolog of S. pombe CDC21; Cdc54p
                     [Saccharomyces cerevisiae]"
                     /protein_id="AAO52157.1"
                     /db_xref="GI:28829641"
                     /translation="MDPKETKTSKISFLNTTKKNNNNNNNNNNNNNNNNNNNNNNNNN
                     NNNNNNNNNDDNNNNGENNIDIEPDDIDLLNKNDIDNIVDGIDNNNKSGSNDMKGLVF
                     WGTTVKVEDVKERFRKFIFNFPTPKERKRFNINDINNNNNNNNNNNNSEEMNDDDDYYY
                     DEEQDKLIKKPEKVQYLYRDLLKVLNETKKCHLNINMKYLVQFDMELYLQWVSFPNEM
                     IPLLDEEINLIYRELFPITPRNGDDEDDDDDDDEDEEDYRIELHPFNLLRKTPMRDLN
                     PSDIDKIISISGLIIRSSSIIPEIKQAFFMCAVCEATFHANVEKGKIQEPSECSNCKS
                     KQSLSIIHNRCLFGDKQYIKLQETPDAIPEGETPHTVALFAYGDLIDIAKPGDRVELT
                     GVFKASPMRAGSNRSLRSIYKTYIDILHIKRTDKGKYDDDDDHDDNTGGGGTGTGKE
                     TNENLDFEDLDEFDLSEEKEREIIELSKKPDIYDIVTKSIAPNIWELEDIKKGILCQL
                     FGGSKKSYQDYGGKFRGDINILLCGDPGTSKSQLLSYVHKIAPRGIYTSGKGSSAVGL
                     TAYITKDPDTKETVLESGALVLSDKGVCCIDEFDKMNDQTRSILHEVMEQQTVSIAKA
                     GIICTLNARTSILASANPSGSRYMPKLSVVENIQLPPTLLSRFDLIYLVLDKANERSD
                     RQLARHLVSMYWDETPVSHFTIPKETLTNYIQYARKHINPKLTDDSAKCLVQGYLEMR
                     SMGSSKKTISATPRQLESLIRIAEAHARIRFSEFVEPLDVEEAIRLIKVALQQAAIDP
                     ENGTIDMDLITTGRSASSREAITRLKSHIKQKIGKKHLTLDQLLKLLTIQNQVQTIQQ
                     IEEIKEALRQLQDEEIIQSSGGIDPTYGVV"
     CDS             join(8672. .9027,9184. .9463)
                     /note="GeneID exon scores (in order of location ranges):
                     37.11, 23.19 - GSCJ_ID dd_02251"
                     /codon_start=1
                     /product="similar to cyctochrome c1 heme lyase
                     [Caenorhabditis elegans]"
                     /protein_id="AAO52158.1"
                     /db_xref="GI:28829642"
                     /translation="MGDDNTNNTSITKSSPTDAMGGGCPIVHDINPANHMYKPNQNPH
                     PDQTKPLSTERITSTIPRTEKDNWQYPSPQMFFNAMKKKQYEPKEEDMSVVISIHNTV
                     NEKCWEDVLQWENDYKDVCPNPKLKKFKGKATDFSPKAKFLNTFLGYKLPFDRHDWIV
                     DRNGKEVRYVIDFYEGRINKDSGKSIGIYIDVRPAIDDFSSLKDRVLHFFK"
     CDS             complement(join(9650. .13297,13390. .13536))
                     /note="GeneID exon scores (in order of location ranges):
                     468.70, 11.48 - GSCJ_ID dd_02250"
                     /codon_start=1
                     /product="similar to C25A11.4b.p [Caenorhabditis elegans]"
                     /protein_id="AAO52159.1"
                     /db_xref="GI:28829643"
                     /translation="MNPFEMIQNQNIVIKNAIEVGCIIVNIGAQDLMNATPYLVLGVI
                     WQIIKQGLLSKVNQNANEILEILFEEDSVKEVDNTQNQSDEHSAEEILLRWVNYHLEK
                     EGCERRISNFSEDIQDSVVYSHLFHQLVPIEFENLVSTIHSESNLFSRAELITKACEK
                     IGVKCFLTPSDIALGHPKLNLALVASLFNSEAAIANLRKEREEKHRTDQERLERERLE
                     RERIERERLERERLEQERLERERLERERLEKQRQEMERIEQERIELERLERERLEQER
                     LEQERIERERLERERLEQERLEQERLERERLEQERLECERLERERLEQERLEQERLER
                     ERIELEELERIERERLEYEERERIERERIEREREELEEQERLEYERLEQERIEREEME
                     RIERERIEREEFERLEEERIEREREEALRLEAERLERELRERERITRQSELERIEYNK
                     EYKICIRIQKRMIDDTSLTDNMTKLIGDITELKRNIIGELRNTTHHQAELKKLTKSID
                     RVLENKNEGKKLKKQKKVIIKPVKGVPNGLDPNKMKNYQSLFYYLQTRPEVMGKLLFL
                     LDRNQMGKTFDNLLLSMFPYEISPREEDLFLKALTVGIEYQIKYSETLQDFLQSESVP
                     MKLLYQYFKKKGFKYLKEIITPLILKVLDQTDLNLNFDPYWIQKEIQIAKELKDGKKS
                     EKKTLDQTFEQLMANNDVVSIFNSRCLQLVELGNAFLERITKSIEQLPVQLRYLCRII
                     AEKSKYAFPTNNTNNHHNGGGISSDESGSTTANITLNNNRSKQLPENREKYISYFLII
                     YRFIIPIISSPDLLNELVLLQHPYITKQSRFNLSQLSRLIQSVFSEKPWMNYFPTVYL
                     NKWIIQNSSSADKFLCQVAQVPDLDELIHYSEYNEIMLENNIKFCSVINISEIIWLHE
                     TILSNFYDLEDIGTLKEVSEKSIQNLAERAINSKLDKGKYVTAARQSTPAIMTSQKSQ
                     NWPLKKYLDEIDFPIQLPEQGDRSIQLSFINRLDQLNEQSIDLWKECKLWMIDILSKS
                     QPPKKNANTIEIIREAIDYFKTGSNSSENLSIINKSNELISRLRTFDRDQCTDGLDGG
                     GPDNFNLFLLDVEKELATILRKNKVMRSELTRLHSLLKSIRIDAAPLEHLTKFQNEQL
                     EKAIVKKQLKTTQLTPLPKDKSTIKPHKFILLDLVKKGMILRCELPESSFSKTKIYIK
                     SLGPGLFEIEAKFGPITKSLQLELDDLLEKSKNHIDEFPLEGITLDVTLTIHLLMKLF
                     SL"
     CDS             complement(join(13685. .13812,14033. .14207))
                     /note="GeneID exon scores (in order of location ranges):
                     6.21, 14.57 - GSCJ_ID dd_02249"
                     /codon_start=1
                     /product="hypothetical protein"
                     /protein_id="AAO52160.1"
                     /db_xref="GI:28829644"
                     /translation="MTYTDSRYNTLNRGSKKMDPPFITQEVLASAKISKKDNILINAN
                     AEDGSVHTFSEEEKVAYSKFITERLIDQEDVLSNYVPIDPSTDQLFVNCTDGVLLK"
     CDS             16650. .17525
                     /note="GeneID exon scores (in order of location ranges):
                     78.29 - GSCJ_ID dd_00829"
                     /codon_start=1
                     /product="hypothetical protein"
                     /protein_id="AAO52161.1"
                     /db_xref="GI:28829645"
                     /translation="METKDKKKDMELLGDRMKSYEDDMKIQIEKNKPFIIRLDGHSFS
                     KFTKNFNKPHDIRIHNAMIETSTVLLKTFMPTCIYTFSDEITMCFPSIDESTLEEGKE
                     IPNLAYSGKVQKLISLSSGLASTVFFKSITNAQYDKDTELNLIKLLETCTPHFDARIF
                     TLPSNQEIVNNLIWRSLVDCKRNSVANLGFAHFTPKQMLGLNNTEVKKKLLEEKSIIY
                     ENEPAWYRFGTYLKKEYYTLTTVSPNEPEKQITAIRSKVNCFSFDITKLSNSLNFISF
                     KTLPEGFNLELIESK"
     CDS             complement(join(17735. .18959,19168. .19640,19750. .20190,
                     20286. .20531))
                     /note="GeneID exon scores (in order of location ranges):
                     130.29, 48.89, 51.03, 20.75 - GSCJ_ID dd_00831"
                     /codon_start=1
                     /product="similar to Dictyostelium discoideum (Slime
                     mold). ALG-2 interacting protein X"
                     /protein_id="AAO52162.1"
                     /db_xref="GI:28829646"
                     /translation="MLSIERKRTEKVDFSKPLTKYIKEQFSKAESDQHETQIATLNGL
                     REDVRNLQERTETSKEMVWKYYSILSSLELRFPISENNVRISFPWTDSYRQRKSTLYS
                     IYFERASVLFNYGSIVSQIASSTNRSNIEGVKKACNQFQLAAGVFNKLREYASLHPEC
                     STSADFSSESLQALVTIMLAQAQECIYEKASMDNLSDSILSKLAAQVAEYYDTFNQLL
                     NSNSLKSIVDRNWNITATVKSYLYKAISLYCHAKGLEQVSQFGEQVSRLMIAVDNINQ
                     SKVNLAKTAPIELKEIVERYVISITRYCESAKKDNDTIYHDTIPPAHKLTPIEKKPLA
                     KALPLPEINFVDPFNSLVPFSVKEDSAYYNDQKETLLRKELDNIEFHNQSAKASLLSM

Alignment Scores:
Pred. No.:              8.39e-61        Length:         153751
Score:                  1909.00         Matches:        397
Percent Similarity:     65.87%          Conservative:   128
Best Local Similarity:  49.81%          Mismatches:     174
Query Match:            41.42%          Indels:         98
```

```
DB:                    2            Gaps:        14
US-10-768-511-8 (1-901) x AC116551 (1-153751)

Qy    179 TrpGlyThrAsnValAsnIleProAspValLeuArgAlaIleArgArgPheLeuHisAsn 198
Db   7589 TGGGGTACAACAGTTAAAGTTGAAGATGTTAAAGAAAGATTTAGAAAATTTATATTTAAT 7530

Qy    199 Tyr--------------------ArgSerSerAlaHisAspLeuAsnSerLys------ 209
Db   7529 TTCCCAACACCAAAAGAAAGAAAAAGATTCAATATAAATGACATAAATAATAATAATAAT 7470

Qy    209 ------------------------------------------------------------ 209
Db   7469 AATAATAACAATAACAATAGTGAAGAAATGAATGATGACGATGATTATTATTATGATGAA 7410

Qy    210 ---------------------------------------------TyrIleGlnIleIle 214
Db   7409 GAACAAGATAAATTAATTAAAAAACCAGAAAAAGTTCAATATTTATATAGAGATTTATTA 7350

Qy    215 GluGluThrValGluArgGluGluAspThrLeuAsnIleAspMetSerAspIleTyrAsp 234
Db   7349 AAAGTATTAAATGAAACTAAAAAATGTCATTTAAATATAAATATGAAATATTTAGTACAA 7290

Qy    235 HisAspProAspLeuTyrAlaLysIleValArgTyrProLeuAspIleIleProLeuLeu 254
Db   7289 TTTGATATGGAATTATATTTACAATGGGTATCATTTCCAAATGAAATGATACCATTATTA 7230

Qy    255 AspThrGluCysGlnGluValAlaThrSerLeuLeuProThrPheGluLysHis------ 272
Db   7229 GATGAGGAAATTAATTTAATTTATAGAGAATTATTTCCAATTACACCAAGAAATGGTGAT 7170

Qy    273 ------------------------------------------------IleGluAlaArg 276
Db   7169 GATGAAGATGATGACGATGATGATGATGAAGATGAAGAGGATTATAGAATTGAATTACAT 7110

Qy    277 ProPheAsnLeuLysAlaSerValHisMetArgGluLeuAsnProSerAspIleAspLys 296
Db   7109 CCATTCAATTTATTAAGAAAAACACCAATGAGAGATTTAAATCCATCAGATATTGATAAG 7050

Qy    297 LeuValSerValLysGlyMetValIleArgCysSerSerIleIleProGluIleLysGly 316
Db   7049 ATTATATCAATTAGTGGTTTAATTATTAGAAGTAGTTCAATTATACCAGAGATTAAACAA 6990

Qy    317 AlaPhePheLysCysLeuValCysGly---------HisSerProProLeuValThrVal 333
Db   6989 GCGTTTTTTATGTGTGCAGTTTGTGAAGCAACTTTTCATGCA------------AATGTT 6942

Qy    334 ValLysGlyArgValGluGluProThrArgCysGluLysProGluCysAlaAlaArgAsn 353
Db   6941 GAGAAGGGTAAAATTCAAGAGCCATCAGAATGTTCA------AATTGTAAATCAAAGCAA 6888

Qy    354 AlaMetSerLeuIleHisAsnArgCysThrPheAlaAsnLysGlnIleValArgLeuGln 373
Db   6887 TCATTATCAATCATTCATAATCGTTGTTTGTTTGGTGATAAACAGTATATTAAACTTCAA 6828

Qy    374 GluThrProAspAlaIleProGluGlyGluThrProHisThrValSerMetCysLeuTyr 393
Db   6827 GAAACACCAGATGCAATTCCAGAGGGTGAAACACCACATACCGTCGCATTATTTGCCTAT 6768

Qy    394 AsnThrMetValAspAlaValLysProGlyAspArgIleGluValThrGlyValPheLys 413
Db   6767 GGTGATTTAATTGATATTGCAAAACCAGGTGATCGTGTCGAGTTAACTGGTGTATTTAAA 6708

Qy    414 AlaMetAlaValArgValGlyProAsnGlnArgThrLeuArgAlaLeuTyrLysThrTyr 433
Db   6707 GCAAGTCCAATGAGAGCTGGTTCAAAT---AGATCACTTAGAAGTATTTATAAAACTTAT 6651

Qy    434 IleAspCysValHisValLysLysSerAspArgGlyArgLeuGlnThrGluAspProMet 453
Db   6650 ATTGATATTTTACATATAAAAAGAACAGATAAAGGTAAATATGATGATGATGATGATGAT 6591

Qy    454 GluMetAsp-----------------------------LysGluAsnAspMetTyrAla 463
Db   6590 CATGATGATAATACTGGTGGTGGTGGCACAGGTACAGGTAAAGAAACAAATGAAAATTTA 6531

Qy    464 GlyTyrHisGluSerAspThrSerGluAlaAlaAsnGluAlaLysIleGlnLysLeuLys 483
Db   6530 GATTTTGAAGATTTAGATGAATTTGATTTATCTGAAGAA---AAAGAAAGAGAGATTATA 6474

Qy    484 GluLeuSerLysLeuProGlyIleTyrAspArgLeuSerArgSerLeuAlaProSerIle 503
Db   6473 GAATTATCAAAAAAACCAGACATTTATGATATTGTAACTAAATCAATAGCACCAAATATT 6414

Qy    504 TrpGluLeuGluAspIleLysLysGlyLeuLeuCysGlnLeuPheGlyGlyLysAlaLys 523
Db   6413 TGGGAATTAGAAGATATTAAAAAAGGTATACTTTGTCAATTATTTGGTGGTTCAAAAAAG 6354

Qy    524 LysIleProSer---GlyAlaSerPheArgGlyAspIleAsnValLeuLeuValGlyAsp 542
Db   6353 AGTTATCAAGATTATGGTGGAAAGTTTAGAGGTGATATAAACATTTTATTGTGTGGTGAT 6294

Qy    543 ProGlyThrSerLysSerGlnLeuLeuGlnTyrValHisLysIleAlaProArgGlyIle 562
Db   6293 CCTGGTACTTCGAAATCTCAATTACTCTCGTATGTTCATAAAATTGCACCCCGTGGTATC 6234

Qy    563 TyrThrSerGlyArgGlySerSerAlaValGlyLeuThrAlaTyrValThrLysAspPro 582
Db   6233 TATACAAGTGGTAAGGGTAGTAGTGCAGTTGGTTTGACGGCATATATCACCAAAGATCCC 6174

Qy    583 GluThrArgGluThrValLeuGluSerGlyAlaLeuValLeuSerAspArgGlyIleCys 602
Db   6173 GATACTAAAGAGACCGTTTTAGAGAGTGGTGCATTGGTACTCTCTGATAAGGGTGTTTGT 6114

Qy    603 CysIleAspGluPheAspLysMetSerAspAsnAlaArgSerMetLeuHisGluValMet 622
Db   6113 TGTATTGATGAATTTGATAAAATGAATGATCAAACTCGTTCAATTCTTCATGAAGTTATG 6054

Qy    623 GluGlnGlnThrValSerValAlaLysGlyGlyIleIleAlaSerLeuAsnAlaArgThr 642
Db   6053 GAACAGCAAACAGTATCAATTGCAAAGGCGGGCATTATCTGCACGTTGAATGCCAGAACC 5994

Qy    643 SerValLeuAlaCysAlaAsnProSerGlySerArgTyrAsnAlaArgLeuSerValIle 662
Db   5993 TCGATTCTTGCATCTGCAAATCCAAGTGGTTCACGTTATATGCCCAAATTATCAGTGGTG 5934

Qy    663 AspAsnIleGlnLeuProProThrLeuLeuSerArgPheAspLeuIleTyrLeuMetLeu 682
Db   5933 GAGAATATCCAATTGCCTCCAACCCTATTATCGAGATTCGATTTAATTTATTTAGTATTG 5874

Qy    683 AspLysProAspGluGlnAsnAspArgArgLeuAlaArgHisLeuValAlaLeuHisTyr 702
Db   5873 GATAAAGCCAATGAAAGAAGTGATCGTCAATTGGCTCGTCATTTAGTATCGATGTATTGG 5814

Qy    703 GluAsnTyrGluValSerLysGlnAspAlaLeuAspLeuGlnThrLeuThrAlaTyrIle 722
Db   5813 GATGAAACACCGGTTTCT---CATTTCACAATCCCCAAGGAAACCCTCACAAACTATATT 5757

Qy    723 ThrTyrAlaArgGlnHisValHisProThrLeuSerAspGluAlaAlaGluAspLeuIle 742
Db   5756 CAATACGCTAGAAAACATATTAATCCAAAGTTAACTGATGATTCTGCAAAATGTTTAGTT 5697

Qy    743 AsnGlyTyrValGluMetArgGlnLysGlyAsnPheProGlySerSerLysLysValIle 762
Db   5696 CAAGGTTATTTAGAGATGAGAAGTATG------------GGTAGTAGTAAAAAAACAATC 5649

Qy    763 ThrAlaThrProArgGlnLeuGluSerMetIleArgIleSerGluAlaLeuAlaArgMet 782
Db   5648 TCTGCCACACCAAGACAATTGGAATCACTTATTCGTATCGCTGAAGCTCATGCTAGAATT 5589

Qy    783 ArgPheSerGluValValGluLysValAspAlaAlaGluAlaValArgLeuLeuAspVal 802
Db   5588 CGTTTCTCTGAATTTGTTGAACCATTAGATGTTGAAGAAGCAATTCGTTTAATTAAAGTC 5529

Qy    803 AlaLeuGlnGlnSerAlaThrAspHisAlaThrGlyThrIleAspMetAspLeuIleThr 822
```

```
Db       5528 GCACTTCAACAAGCTGCAATCGATCCTGAAAATGGTACAATCGATATGGATCTAATTACA 5469

Qy        823 ThrGlyValSerAlaSerGluArgIleArgArgAlaAsnLeuLeuAlaAlaLeuArgGlu 842
              ||||||   |||||||||   |||               |||   :::   :::::::::
Db       5468 ACTGGTCGTTCTGCAAGTTCTCGTGAAGCTATCACTAGATTAAAATCTCACATCAAACAA 5409

Qy        843 LeuIleAlaAspLysIleSerProGlySerSerSerGlyLeuLysThrSerGlnLeuLeu 862
                 |||      |||                        |||         |||||||||
Db       5408 AAAATTGGTAAAAAA---------------------CATTTAACACTCGACCAACTCTTA 5370

Qy        863 GluAspIleArgSerGlnSerSerValAspVal------SerLeuGlnAspIleLysAsn 880
              :::   :::      |||:::   |||               :::::::::||||||
Db       5369 AAATTATTAACAATTCAAAATCAAGTTCAAACTATACAACAAATTGAAGAAATTAAAGAA 5310

Qy        881 AlaLeuGlySerLeuGlnGlyGluGlyPheLeuThrValHisGlyAspIle 897
              ||||||      ||||||   |||      :::         |||   |||
Db       5309 GCTTTAAGACAATTACAAGATGAAGAAATTATTCAATCTTCTGGTGGTATT 5259

RESULT 7
AC007047
LOCUS       AC007047               114418 bp    DNA     linear   PLN 11-MAR-2002
DEFINITION  Arabidopsis thaliana chromosome 2 clone F16F14 map mi398, complete
            sequence.
ACCESSION   AC007047
VERSION     AC007047.7  GI:20198030
KEYWORDS    HTG.
SOURCE      Arabidopsis thaliana (thale cress)
  ORGANISM  Arabidopsis thaliana
            Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
            Spermatophyta; Magnoliophyta; eudicotyledons; core eudicotyledons;
            rosids; eurosids II; Brassicales; Brassicaceae; Arabidopsis.
REFERENCE   1  (bases 1 to 114418)
  AUTHORS   Lin,X., Kaul,S., Shea,T.P., Fujii,C.Y., Shen,M., VanAken,S.E.,
            Barnstead,M.E., Mason,T.M., Bowman,C.L., Ronning,C.M.,
            Benito,M.-I., Carrera,A.J., Creasy,T.H., Buell,C.R., Town,C.D.,
            Nierman,W.C., Fraser,C.M. and Venter,J.C.
  JOURNAL   Unpublished
REFERENCE   2  (bases 1 to 114418)
  AUTHORS   Lin,X.
  TITLE     Direct Submission
  JOURNAL   Submitted (09-MAR-2000) The Institute for Genomic Research, 9712
            Medical Center Dr., Rockville, MD 20850, USA
REFERENCE   3  (bases 1 to 114418)
  AUTHORS   Town,C.D. and Kaul,S.
  TITLE     Direct Submission
  JOURNAL   Submitted (27-FEB-2002) The Institute for Genomic Research, 9712
            Medical Center Dr, Rockville, MD 20850, USA, cdtown@tigr.org
COMMENT     On Apr 18, 2002 this sequence version replaced gi:6598653.
            On Aug 22, 1997 this sequence version replaced gi:1808697.
             The sequence of BAC F19P19 from Arabidopsis thaliana chromosome 1.
FEATURES             Location/Qualifiers
     source          1. .114418
                     /organism="Arabidopsis thaliana"
                     /mol_type="genomic DNA"
                     /db_xref="taxon:3702"
                     /chromosome="2"
                     /map="mi398"
                     /clone="F16F14"
                     /ecotype="Columbia"
     misc_feature    1. .766
                     /note="overlap with BAC clone F1P15
                     (AC007195:55519. .56284)."
     gene            567. .2420
                     /gene="At2g16485"
                     /note="synonym: F16F14.1"
     mRNA            <567. .>2420
                     /gene="At2g16485"
     CDS             567. .2420
                     /gene="At2g16485"
                     /codon_start=1
                     /product="hypothetical protein"
                     /protein_id="AAM15362.1"
                     /db_xref="GI:20198031"
                     /translation="MDSDSERASLESIKDNSECVHVSNEPNLTATCVDSSVGEEGVTD
                     VNSSAAVSELVPPEQGEGALLNSVPEISERGIPVDVVSSVDGGGEENAAFNIQEIDSV
                     GGDAAAVEEVPLKSSSVVGEGREEEAGASIVKEEDFVAEANLSGDRLEENKEVSMEEE
                     PSSHELSVCEVNGVDSLNDEENREVGEQIVCGSMGGEEIESDLESKKEKVDVIEEETT
                     AQAASLVNAIEIPDDKEVACVAGFTEISSQDKGLDESGNGFLDEEPVKELQIGEGAKD
                     LTDGDAKEGVDVTEDEMDIQVLKKSKEEEKVDSTTELEIETMRLEVHDVATEMSDKTV
                     ISSAVVTQFTGETSNDKETVMDDVKEDVDKDSEAGKSLDIHVPEATEEVDTDVNYGVG
                     IEKEGDGVGGAEEAGQTVDLEEIREENQELSKELAQVDETKISEMSEVTETMIKDEDQ
                     EKDDNMTDLAEDVENHRDSSVADIEEGREDHEDMGVTETQKETVLGKVDRTKIAEVSE
                     ETDTRIEDEDQEKDDEMTDVAEDVKTHGDSSVADIEEGRESQEEMTETQEDSVMADEE
                     PEEVEEENKSAGGKRKRGRNTKTVKGTGKKKEEDVCFMCFDGGDLVLCDRR"
     gene            3940. .5514
                     /gene="At2g16480"
                     /note="synonym: F16F14.2; predicted by genscan and
                     genefinder"
     mRNA            join(<3940. .4068,4408. .5287,5372. .>5514)
                     /gene="At2g16480"
     CDS             join(3940. .4068,4408. .5287,5372. .5514)
                     /gene="At2g16480"
                     /codon_start=1
                     /product="hypothetical protein"
                     /protein_id="AAD22293.1"
                     /db_xref="GI:4544383"
                     /translation="MFSLCKGCAKDAVFFCIRGNKGLCETCMETVKLIERKQQEKEPA
                     QLDFNDKTSWEYLFKDYWIDLKTQLSLSPEELDQAKRPLKGHETNASKQGTASETDYV
                     TDGGSDSDSSPKKRKTRSRSKSGSAEKILSSGDKNLSDETMEWASKELLDLVVHMRRG
                     DRSFLPMLEVQTLLLAYIKRYNLRDPRRKSQVICDSRLQNLFGKSHVGHFEMLNLLDS
                     HFLKKEQNQADDIQGDIVDTEEPNHVDVDENLDHPVKSGKDKKRKTRKKNVRKGRQSN
                     LDDFAAVDMHNINLIYLRRSLVEDLLEDSTAFEEKVASAFVRLRISGNQKQDLYRLVQ
                     VVGTSKAPEPYKVGKKTTDYVLEILNLDKTEVISIDIISNQDFTEVTYC"
     gene            6424. .9042
                     /gene="At2g16470"
                     /note="synonym: F16F14.3"
     mRNA            join(<6424. .6583,6701. .8419,8909. .>9042)
                     /gene="At2g16470"
     CDS             join(6424. .6583,6701. .8419,8909. .9042)
                     /gene="At2g16470"
                     /codon_start=1
                     /product="unknown protein"
                     /protein_id="AAD22314.1"
                     /db_xref="GI:4544404"
                     /translation="MLTLRECVEKLQLLKSPEERQRRLEEIPEIHADPKMDPDCESED
                     EDEKEEKEKEKQLRPRSSSFNRRGRDPISPRKGGFSSNESWTGTSNYSNTSANRELSR
                     SYSGRGSTGRGDYLGSSDDKVSDSMWTSAREREVQPSLGSEKPRSVSIPETPARSSRA
                     IAPPELSPRIASEISMAPPAVVSQPVPKSNDSEKIWHYKDPSGKVQGPFSMAQLRKWN
                     NTGYFPAKLEIWKANESPLDSVLLTDALAGLFQKQTQAVDNSYMKAQVAAFSGQSSQS
                     EPNLGFAARIAPTTIEIPRNSQDTWSQGGSLPSPTPNQITTPTAKRRNFESRWSPTKP
                     SPQSANQSMNYSVAQSGQSQTSRIDIPVVVNSAGALQPQTYPIPTPDPINVSVNHSAT
                     LHSPTPAGGKQSWGSMQTDHGGSNTPSSQNNSTSYGTPSPSVLPSQSQPGFPPSDSWK
                     VAVPSQPNAQAQAQWGMNMVNNNQNSAQPQAPANQNSSWGQGTVNPNMGWVGPAQTGV
                     NVNWGGSSVPSTVQGITHSGWVAPVQGQTQAYPNPGWGPTGHPQSQSQSQVQAQAGTT
                     GSGWMQPGQGIQSGNSNQNWGTQNQTAIPSGGSGGNQAGYWGNQQQSQNGDSGYGWNR
                     QSGLGKNIGTDYLKDPYPIGFYLNPTRLPECPPLVPLTQIHVVNES"
     repeat_region   complement(8530. .8554)
                     /rpt_family="(GA)n"
     repeat_region   complement(9636. .9668)
                     /rpt_family="AT_rich"
     gene            10142. .12397
                     /gene="At2g16460"
                     /note="synonym: F16F14.4; supported by full length cDNA:
                     Ceres:105900"
     mRNA            join(10142. .10485,10634. .10765,10848. .10907,11485. .11561,
                     11641. .11697,11809. .11875,11966. .12397)
                     /gene="At2g16460"
     CDS             join(10306. .10485,10634. .10765,10848. .10907,11485. .11561,
                     11641. .11697,11809. .11875,11966. .12085)
                     /gene="At2g16460"
                     /codon_start=1
                     /product="expressed protein"
                     /protein_id="AAD22294.2"
                     /db_xref="GI:20198032"
                     /translation="MTLSKRLALLGAQSAITLAKPRGLGSSLGLLDRRPFPYRQFSEL
                     TKANGRRAFLVDTLALVRSLEAQGVPSKQAEAITSAITEVLNDSLENVSESFVSKAEM
```

```
                     QKIEMIQDSNLSKFKSEVKSSQEHHFTVLQRETEKLRGDIEKMRSELRYEIDKVTAGQ
                     RLDLNLERGRIRDELANQNTETTNLTNKLDREIHALRAQLEASKYEVIKYCIGTLVSI
                     SAVGLAVLRIMI"
     gene            12986. .14269
                     /gene="At2g16450"
                     /note="synonym: F16F14.5; predicted by genscan and
                     genefinder"
     mRNA            <12986. .>14269
                     /gene="At2g16450"
     CDS             12986. .14269
                     /gene="At2g16450"
                     /codon_start=1
                     /product="hypothetical protein"
                     /protein_id="AAD22295.1"
                     /db_xref="GI:4544385"
                     /translation="MNPSPITIDLILEILSRLPAKSVRRFHCVSKRWASIFGSPYFKE
                     LFLTRSSTKPRLLFAIAEKGNKEKDCVWRFFSSPQLENPYEKSSSTLVATAEFHVRFS
                     PDNLLICHYYDLKYFSIGYAFGLIYIYGNRGRARPLICNPTTGRYAILPNRYTYRKAF
                     SFFGFDPIDKQYKALSMVYPSGPGHSRVITFGAGDLKWRRIKCSLRHDIKSEGVCING
                     VLYYLGDTSDWSRVNGNHVTSGYMIVCFDVRSEKFTFIDVKRFCRLINYKGKLAVIYW
                     EDDVDIQELYYKKGIDVEEYVENNVNADATNELCVWILADVEKQEWSKHAYTWTDEKF
                     FRRLVSIAGVTASGEIVFSMRKCNPKQPFYVFYFNPERNSLQRVEIQGFGEAVTKSCD
                     VCTFVNHVEDLNVYDLKQLKSVHPPLVEPEYYDSD"
     gene            15483. .19098
                     /gene="At2g16440"
                     /note="synonym: F16F14.6"
     mRNA            join(<15483. .15918,16016. .16144,16218. .16299,16372. .16504,
                     16588. .16659,16757. .16858,16971. .17084,17164. .17376,
                     17485. .17583,17709. .17869,17969. .18065,18166. .18327,
                     18429. .18509,18583. .18661,18771. .18873,18999. .>19098)
                     /gene="At2g16440"
     CDS             join(15483. .15918,16016. .16144,16218. .16299,16372. .16504,
                     16588. .16659,16757. .16858,16971. .17084,17164. .17376,
                     17485. .17583,17709. .17869,17969. .18065,18166. .18327,
                     18429. .18509,18583. .18661,18771. .18873,18999. .19098)
                     /gene="At2g16440"
                     /codon_start=1
                     /product="putative CDC21 protein"
                     /protein_id="AAD22296.1"
                     /db_xref="GI:4544386"
                     /translation="MSTDEPLPSSDDGEEDGGDDTTPTFVWGTNISVQDVKSAIEMFV
                     KHFREARENSDDLFREGKYMVSIRKVIEIEGEWIDVDAFDVFDYDPDLYNKMVRYPLE
                     VLAIFDIVLMDIVSTINRLFEKHVQVRIFNLRTSTSMRNLNPSDIEKMISLKGMIIRS
                     SSIIPEIREAVFRCLVCGYFSDPIIVDRGKISEPPTCLKQECMTKNSMTLVHNRCRFA
                     DKQIVRLQETPDEIPEGGTPHTVSLLLHDKLVDNGKPGDRIEVTGIYRAMTVRVGPAH
                     RTVKSVFKTYIDCLHIKKASKLRMSAEDPMDVDNSLRRVDEDLRKFQELSKQPDIYER
                     LSRSLAPNIWELDDVKKGLLCQLFGGNALNLASGANFRGDINILLVGDPGTSKSQLLQ
                     YIHKLSPRGIYTSGRGSSAVGLTAYVAKDPETGETVLESGALVLSDRGICCIDEFDKM
                     SDSARSMLHEVMEQQTVSIAKAGIIASLNARTSVLACANPSGSRYNPRLSVIENIHLP
                     PTLLSRFDLIYLILDKPDEQTDRRLAKHIVALHFENAESAQEEAIDITTLTTYVSYAR
                     KNIHPKLSDEAAEELTRGYVELRKAGKFAGSSKKVITATPRQIESLIRLSEALARMRF
                     SEWVEKHDVDEAFRLLRVAMQQSATDHATGTIDMDLINTGVSASERMRRDTFASSIRD
                     IALEKITMEMLLLQLLEELKKHGGNINTEIHLHDVSHL"
     repeat_region   complement(20438. .20463)
                     /rpt_family="(TA)n"
     repeat_region   complement(20932. .20956)
                     /rpt_family="AT_rich"
     repeat_region   complement(21489. .21518)
                     /rpt_family="AT_rich"
     gene            23017. .25287
                     /gene="At2g16430"
                     /note="synonym: F16F14.7"
     mRNA            join(<23017. .23199,23497. .23697,23793. .23904,24010. .24239,
                     24350. .24488,24580. .24800,24878. .24999,25089. .>25287)
                     /gene="At2g16430"
     CDS             join(23017. .23199,23497. .23697,23793. .23904,24010. .24239,
                     24350. .24488,24580. .24800,24878. .24999,25089. .25287)
                     /gene="At2g16430"
                     /codon_start=1
                     /product="putative purple acid phosphatase precursor"
                     /protein_id="AAD22297.1"
                     /db_xref="GI:4544387"
                     /translation="MGRVRKSDFGSIVLVLCCVLNSLLCNGGITSRYVRKLEATVDMP

                     LDSDVFRVPCGYNAPQQVHITQGDVEGKAVIVSWVTQEAKGSNKVIYWKENSTKKHKA

Alignment Scores:
Pred. No.:              5.79e-60        Length:         114418
Score:                  1883.00         Matches:        498
Percent Similarity:     45.87%          Conservative:   146
Best Local Similarity:  35.47%          Mismatches:     190
Query Match:            40.85%          Indels:         573
DB:                     15              Gaps:           30

US-10-768-511-8 (1-901) x AC007047 (1-114418)

Qy         13 SerSerProTyrProSerGlnSerGluGlyValSerThrProLeuProGlnValThrSer 32
                |||   |||   ||||||       |||   ||||||:::|||:::          |||
Db      15225 TCAGGTCCT---CCTTCTCCCGGAGAAAACGTTTCAAGCCCTATCGAAAACACATACTCT 15281

Qy         33 ProSerPheAspAsnAlaAlaSerProValAlaGlyArgArgAlaValArgGlnThrPro 52
                                  ||||||   |||      |||
Db      15282 --------------------TCCCCCGCCGCACTTCACCGT------------------ 15302

Qy         53 ThrSerAlaValArgArgArgGlyArgGluThrAspSerAlaArgArgArgArgSerArg 72
                           |||||||||||||||
Db      15303 -----------CGCCGGCGAGGAAGATCC------------------------------ 15320

Qy         73 SerArgSerLeuGlyAsnSerValTyrSerSerProTyrAspAlaGlyThrProGlyThr 92
                                         |||:::|||               ||||||
Db      15321 --------------------------TCTACTCCCACGCAATTCGCAACTCCTCCACCA 15353

Qy         93 ProGlyThrProValAlaThrProValTyrAlaThrProValGlyThrProMetGlyThr 112
              |||   :::   :::|||:::             :::||||||        |||
Db      15354 CCGCCTTCACGTCTCGCATCATCC---AATTCCACTCCT---------CCAACATCTCGT 15401

Qy        113 ProSerPheHisArgGlyThrProGlnTyrLysGlnArgSerGluLeuGlySerGlnGly 132
              ||||||      |||                     |||   |||:::        |||
Db      15402 CCTTCCGCCGCCAGATCT------------AAAGGAAGAAATGGACACGGAGGTGGCGGT 15449

Qy        133 LysProLeuHisArgArgArgArgSerGlnSerArgGluProGlyHisArgSerProSer 152
                                                :::||||||             :::
Db      15450 -----------------GGTGGCGGTGGTGGCGGAGACCCAGGGACGCCAATGTCCACT 15491

Qy        153 ArgGluProSerAlaAspGlyArgProSerGluSerAlaGluProAspAspThrLeuGly 172
                 ||||||                        |||         |||||||||
Db      15492 GACGAGCCACTGCCTTCTTCCGATGACGGTGAAGAAGACGGAGGCGACGATACTACTCCG 15551

Qy        173 GlyGluTyrAlaTyrValTrpGlyThrAsnValAsnIleProAspValLeuArgAlaIle 192
                       :::|||||||||||||||:::::::::      ||||||        ||||||
Db      15552 ---------ACTTTCGTTTGGGGAACTAACATCAGTGTGCAAGATGTTAAATCTGCGATT 15602

Qy        193 ArgArgPheLeuHisAsnTyr---------ArgSerSerAlaHisAspLeu------Asn 207
                    |||:::   ::::::            |||      ::::::   ||||||
Db      15603 GAAATGTTTGTTAAGCATTTCAGAGAAGCTAGAGAGAATAGTGATGATTTGTTTAGAGAA 15662

Qy        208 SerLysTyrIleGlnIleIleGluGluThrValGluArgGluGluAspThrLeuAsnIle 227
                 ||||||:::        |||   :::   :::|||   |||   :::   :::::::::
Db      15663 GGGAAGTATATGGTTTCGATAAGGAAGGTGATTGAGATTGAAGGAGAATGGATTGATGTT 15722

Qy        228 AspMetSerAspIleTyrAspHisAspProAspLeuTyrAlaLysIleValArgTyrPro 247
              |||      |||:::::::|||:::||||||||||||||||    |||:::||||||||||||
Db      15723 GATGCGTTTGATGTGTTTGATTATGATCCTGATTTGTATAACAAGATGGTTCGTTATCCT 15782

Qy        248 LeuAspIleIleProLeuLeuAspThrGluCysGlnGluValAlaThrSerLeuLeuPro 267
              |||:::::::::    :::   |||          ::::::   :::::::::
Db      15783 TTGGAAGTTCTCGCTATCTTCGATATTGTGCTGATGGATATTGTTTCCACGATTAATCGT 15842

Qy        268 ThrPheGluLysHisIleGluAlaArgProPheAsnLeuLysAlaSerValHisMetArg 287
                 |||||||||||||:::::   |||   |||||||||:::   |||      ||||||
Db      15843 TTGTTCGAGAAACATGTTCAAGTTAGGATTTTTAACCTCAGGACATCTACTTCCATGAGA 15902

Qy        288 GluLeuAsnProSer------------------------------------------- 292
                 ||||||||||||
Db      15903 AATCTCAATCCATCTGGTTAGTATGATTCAGTTTTTGTTCTTCTTGGTTTCTTAGTTCTC 15962

Qy        293 ---------------------------------------------------AspIleAs 295
```

```
Db      15963 TCTAGTGGTAATAGCTTAGCTTATGTTTGGATGTGTTTTGTTTGGTTGGTTAGATATCGA 16022
Qy        295 pLysLeuValSerValLysGlyMetValIleArgCysSerSerIleIleProGluIleLy 315
Db      16023 GAAGATGATCTCTTTGAAAGGTATGATAATTCGGAGTAGCTCTATCATTCCTGAAATAAG 16082
Qy        315 sGlyAlaPhePheLysCysLeuValCysGlyHis---SerProProLeuValThr----- 332
Db      16083 GGAAGCTGTGTTTAGATGCCTTGTTTGTGGTTACTTCTCTGACCCCATCATCGTTGATAG 16142
Qy        332 ------------------------------------------------------------ 332
Db      16143 AGGTAATTTGGTCATTTGGATTTATCCAGATTTTATGTATCTGCATAGTTTCTCATAATG 16202
Qy        333 ----ValValLys-GlyArgValGluGluProThrArgCysGluLysProGluCysAlaA 351
Db      16203 TTGTTTGATTGATAGGGAAAATAAGTGAGCCTCCTACTTGCTTGAAACAAGAGTGCATGA 16262
Qy        351 laArgAsnAlaMetSerLeuIleHisAsnArgCys------------------------- 362
Db      16263 CCAAGAACTCCATGACACTAGTGCACAACCGATGCAGGTAAGTTCTAGACATAGTATATG 16322
Qy        363 -----------------------------------------------ThrPheAlaAsnL 367
Db      16323 TCTGATTCTTATAATTTCTGTAACTTACCTTCTTTATTTTGTGTATCAGGTTTGCTGATA 16382
Qy        367 ysGlnIleValArgLeuGlnGluThrProAspAlaIleProGluGlyGluThrProHisT 387
Db      16383 AGCAGATTGTGAGGCTTCAGGAAACACCTGATGAGATTCCAGAAGGAGGAACACCTCACA 16442
Qy        387 hrValSerMetCysLeuTyrAsnThrMetValAspAlaValLysProGlyAspArgIleG 407
Db      16443 CAGTTAGCTTGTTGCTGCATGATAAGCTAGTTGATAATGGGAAGCCTGGTGATAGAATTG 16502
Qy        407 luVal------------------------------------------------------- 408
Db      16503 AGGT-AATAAAAGCAGTAATTCTATCAGGGTAGTACATTCATTCCTATGGTTTTCGTTGA 16561
Qy        409 -----------------------------ThrGlyValPheLysAlaMetAlaValArgV 419
Db      16562 CAATCAAAAGCTTGTTTTCTGGATAGGTCACCGGAATTTACAGAGCAATGACTGTCCGAG 16621
Qy        419 alGlyProAsnGlnArgThrLeuArgAlaLeuTyrLys---------------------- 431
Db      16622 TGGGACCTGCTCACAGGACTGTGAAATCTGTGTTTAAGGCATGTCATCTTATTTCTTCCT 16681
Qy        431 ------------------------------------------------------------ 431
Db      16682 CTGTTAGTTATTTCCCTTTATTGTAATTACGTTCTGACATTGTCATTTATGTTTTCTTAT 16741
Qy        432 ---------------ThrTyrIleAspCysValHisValLysLysSerAspArgGlyArg 446
Db      16742 GAATGTTGAGGACAGACCTATATCGATTGCCTTCACATAAAGAAAGCAAGTAAGTTAAGA 16801
Qy        447 LeuGlnThrGluAspProMetGluMetAspLys--------------------------- 457
Db      16802 ATGTCTGCTGAGGATCCTATGGATGTTGACAATAGTCTGCGTAGAGTTGATGAAGATGTC 16861
Qy        458 GluAsnAsp---------------------------------------------MetTyr 462
Db      16862 GAGTTAGATGAGGAGAAGGTATTTCATCTATTTAACTAAAATGGTTCTACAGTTTTGTAT 16921
Qy        463 AlaGlyTyr-HisGluSerAspThrSerGluAlaAlaAsnGluAlaLysIleGlnLysLe 482
Db      16922 CTTATTCATTCGTCCCTCCATTCCAACCAATATTGTGTTACTTTTCAGCTGAGGAAGTT 16981
Qy        482 uLysGluLeuSerLysLeuProGlyIleTyrAspArgLeuSerArgSerLeuAlaProSe 502
Db      16982 TCAGGAACTCTCTAAACAACCAGATATATATGAGAGGCTTAGTAGATCACTAGCACCAAA 17041
Qy        502 rIleTrpGluLeuGluAspIleLysLysGlyLeuLeuCysGln----------------- 516

Db      17042 CATCTGGGAGTTGGATGATGTTAAAAAGGGTCTTCTATGCCAGGTACTGTAACGCTTTTG 17101
Qy        516 ------------------------------------------------------------ 516
Db      17102 ATACTGACTTATGGTTTAAACAAAATTAGTTCTTTTTACTGAGTCAAATTTCAAATGTGA 17161
Qy        517 --LeuPheGlyGlyLysAlaLysLysIleProSerGlyAlaSerPheArgGlyAspIleA 536
Db      17162 AGCTTTTTGGAGGGAATGCTTTGAACTTAGCATCTGGTGCTAATTTTCGTGGTGACATCA 17221
Qy        536 snValLeuLeuValGlyAspProGlyThrSerLysSerGlnLeuLeuGlnTyrValHisL 556
Db      17222 ACATCTTACTTGTTGGTGACCCTGGTACGAGCAAGTCCCAGCTGCTCCAGTACATTCATA 17281
Qy        556 ysIleAlaProArgGlyIleTyrThrSerGlyArgGlySerSerAlaValGlyLeuThrA 576
Db      17282 AGCTTTCACCCCGTGGCATTTACACAAGTGGGCGAGGGAGCTCAGCTGTTGGTTTAACAG 17341
Qy        576 laTyrValThrLysAspProGluThrArgGluThr------------------------- 587
Db      17342 CTTATGTAGCTAAAGATCCTGAGACAGGAGAAACTGTATGCTAATATTGTTTTCAATCTT 17401
Qy        587 ------------------------------------------------------------ 587
Db      17402 TTTATTATCAGCTGTTTTTTTTTATTCTGTGTTTCAAAGTTTTCATAAACCTGACATGCT 17461
Qy        588 -----------------------ValLeuGluSerGlyAlaLeuValLeuSerAspArgG 600
Db      17462 AACATGAATGCTACATTGTTTAGGTTTTGGAGAGTGGAGCTCTTGTTCTTAGTGACCGAG 17521
Qy        600 lyIleCysCysIleAspGluPheAspLysMetSerAspAsnAlaArgSerMetLeuHisG 620
Db      17522 GTATCTGCTGTATTGATGAATTTGACAAAATGTCTGACAGTGCGAGGAGCATGCTACACG 17581
Qy        620 luVal------------------------------------------------------- 621
Db      17582 AGGT-TTGTTCCTCTCTCTCTTTTTTAGCATTGCTCAAAGTCGTAGCTTACAGTTTTTGG 17640
Qy        621 ------------------------------------------------------------ 621
Db      17641 GTGCCCCCTGTGAATGTGGTCTCTAGTCAATAAGATTGCATTACTGATGAACATTGTTTC 17700
Qy        622 -----------MetGluGlnGlnThrValSerValAlaLysGlyGlyIleIleAlaSerL 638
Db      17701 TCTGTCAGGTGATGGAACAGCAGACTGTTTCAATAGCAAAGGCTGGTATTATTGCATCTC 17760
Qy        638 euAsnAlaArgThrSerValLeuAlaCysAlaAsnProSerGlySerArgTyrAsnAlaA 658
Db      17761 TAAATGCCAGAACCTCTGTGTTGGCTTGTGCAAATCCTAGTGGCTCACGTTATAATCCCC 17820
Qy        658 rgLeuSerValIleAspAsnIleGlnLeuProProThrLeuLeuSer------------- 673
Db      17821 GGCTTTCTGTTATTGAGAATATTCACCTTCCTCCAACCTTGCTATCTAGGTATGTTGCAA 17880
Qy        673 ------------------------------------------------------------ 673
Db      17881 TCCTTCAACAATGTGACATTATTTTTACTAAAGGCTAATGGTTTCCTCTATTTGGTGTTA 17940
Qy        674 --------------------------ArgPheAspLeuIleTyrLeuMetLeuAspLysP 685
Db      17941 CTAATTATAATCTCATTTTGCTTCTTAGATTCGATTTGATCTACTTGATTCTTGACAAGC 18000
Qy        685 roAspGluGlnAsnAspArgArgLeuAlaArgHisLeuValAlaLeuHisTyrGluAsn- 704
Db      18001 CCGATGAGCAGACTGACCGAAGGCTTGCAAAGCATATTGTGGCCCTTCACTTTGAGAACG 18060
Qy        704 ------------------------------------------------------------ 704
Db      18061 CTGAGGTCTTATATTATATATTTCCCAGTGTGTATGATATGTTGACAGACAATTATTTT 18120
Qy        705 --------------------------------------Tyr-GluValSerLysGlnAsp 711
Db      18121 TTGCCTTTATCTTACTTAATTGGTATCATTTGTGTCAATATACAGAGTGCTCAGGAGGAA 18180
```

```
Qy      712 AlaLeuAspLeuGlnThrLeuThrAlaTyrIleThrTyrAlaArgGlnHisValHisPro 731
            |||:::|||:::    |||||||||     |||::::::||||||||:::::::::||||||
Db    18181 GCTATAGATATTACTACACTGACAACTTATGTTAGCTATGCCCGCAAGAACATTCATCCT 18240

Qy      732 ThrLeuSerAspGluAlaAlaGluAspLeuIleAsnGlyTyrValGluMetArgGlnLys 751
                |||||||||||||||||||:::|||       ||||||||||:::|||:::
Db    18241 AAACTGTCAGATGAAGCTGCAGAGGAGTTGACCCGAGGATATGTTGAGCTGAGAAAAGCA 18300

Qy      752 GlyAsnPheProGlySerSerLysLysVal------------------------------ 761
            |||   |||   ||||||||||||||
Db    18301 GGGAAGTTTGCTGGCAGTAGCAAAAAGGT-GTGCCCTCGAAACTTCAATATCTTTTCTAT 18359

Qy      761 ------------------------------------------------------------ 761

Db    18360 GTCGTAACCAAGGCTCATAAAAAGGCTATGACTAGTCTAACGCTATGAAACGCTCTGAAT 18419

Qy      762 ------------IleThrAlaThrProArgGlnLeuGluSerMetIleArgIleSerGlu 777
                        ||||||||||||||||||||:::||||||:::||||||:::||||||
Db    18420 TTGGCACAGGTCATAACAGCAACTCCTAGGCAGATTGAAAGTTTGATCAGACTTAGTGAG 18479

Qy      778 AlaLeuAlaArgMetArgPheSerGlu--------------------------------- 786
            |||||||||||||||||||||||||||
Db    18480 GCCTTGGCTCGGATGCGCTTCTCCGAATGGGTTAGTTAAACGCTCATATTATACATTTTC 18539

Qy      787 ------------------------------------------ValValGluLysValAspAl 793
                                                      ||||||||   |||
Db    18540 CTTTCATCGTTCTAAATCAACTGAAATGGTTTTTGTGATAAAGGTTGAAAAGCATGACGT 18599

Qy      793 aAlaGluAlaValArgLeuLeuAspValAlaLeuGlnGlnSerAlaThrAspHisAlaTh 813
                 ||||||    |||||||||   ||||||:::|||||||||||||||||||||||||
Db    18600 GGACGAGGCATTTCGACTTCTCAGAGTTGCAATGCAGCAATCAGCAACGGATCACGCCAC 18659

Qy      813 r----------------------------------------------------------- 813
            |
Db    18660 AGGTAGTTTTTTCACAGTAACTTTCAGACCCATTATTATCATCAGGATAAAGATAAACAG 18719

Qy      814 ---------------------------------------------------GlyThrIleA 817
                                                               |||||||||
Db    18720 TAGCATTCAATACCAATCTAACAATAAACACCTGTATTTATGTTGGTGAAGGGACCATTG 18779

Qy      817 spMetAspLeuIleThrThrGlyValSerAlaSerGluArgIleArgArgAlaAsnLeuL 837
             ||||||||||||    |||||||||||||||||||||||:::||||||
Db    18780 ACATGGACCTGATAAATACTGGAGTGTCAGCAAGTGAACGAATGAGGCGGGATACCTTTG 18839

Qy      837 euAlaAlaLeuArgGluLeuIleAlaAspLysIleSerProGlySerSerSerGlyLeuL 857
              ::::::::|||::::::      :::|||:::       |||    ||||||    ::::
Db    18840 CGTCGTCAATTAGAGACATAGCTCTGGAGAAGATGCAAATTGGAGGATCATCA---ATGC 18896

Qy      857 ysThrSerGlnLeu---------------------------------------------- 861
            ::    |||::::::
Db    18897 GCTTATCCGAGGTATTCATGCACTACTCCTTTCTTTTTCCCAATTTTGTTATATGCCAAA 18956

Qy      861 ------------------------------------------------------------ 861

Db    18957 AGCAAGCAAATCAGTTTACAATAAGAGAAACCGTTTCTTAAGTACAATGGAAATGTTGTT 19016

Qy      862 -----------LeuGluAspIleArgSerGlnSerSer---------------------- 870
                       |||::::::   ||||||      :::
Db    19017 GTTGCAGTTACTTGAAGAACTGAAGAAGCATGGAGGCAACATAAACACCGAAATTCATCT 19076

Qy      870 ------------------------------------------------------------ 870

Db    19077 TCATGATGTGAGTCATTTATAACACTTTGCTACATATCAAAAGGTTTTTGGTCTAAGGAC 19136

Qy      871 --------------------ValAspValSer-LeuGlnAspIleLysAsnAlaLeuGly 883
                                |||   |||::: ::::::   ::::::   |||:::
Db    19137 AGAGACATAACAAAGTATTTGTTGTTGTGACGCATTAAACAGGTTAGAAAAGCAGTCGCG 19196

Qy      884 SerLeuGlnGlyGluGlyPheLeuThrValHisGlyAspIleValLysArgVal 901
            :::|||         ||||||||||||            ||||||   :::|||||||||
Db    19197 ACGCTAGCGAGTGAAGGATTTCTGGTAGCTGAAGGTGATAGAATCAAGAGAGTA 19250
```

```
RESULT 8
SPCDC21
LOCUS       SPCDC21                 3667 bp    DNA     linear   PLN 02-MAR-2000
DEFINITION  Schizosaccharomyces pombe cdc21 gene for protein required for S
            phase execution.
ACCESSION   X58824
VERSION     X58824.1  GI:4165292
KEYWORDS    cdc21+ gene; cell division control; mcm4.
SOURCE      Schizosaccharomyces pombe (fission yeast)
  ORGANISM  Schizosaccharomyces pombe
            Eukaryota; Fungi; Ascomycota; Schizosaccharomycetes;
            Schizosaccharomycetales; Schizosaccharomycetaceae;
            Schizosaccharomyces.
REFERENCE   1
  AUTHORS   Coxon,A., Maundrell,K. and Kearsey,S.E.
  TITLE     Fission yeast cdc21+ belongs to a family of proteins involved in an
            early step of chromosome replication
  JOURNAL   Nucleic Acids Res. 20 (21), 5571-5577 (1992)
   PUBMED   1454522
REFERENCE   2
  AUTHORS   Kearsey,S.E.
  TITLE     Direct Submission
  JOURNAL   Submitted (05-APR-1991) S.E. Kearsey, Oxford University, South
            Parks Road, Oxford, UK
  REMARK    revised by [3]
REFERENCE   3  (bases 1 to 3667)
  AUTHORS   Kearsey,S.E.
  TITLE     Direct Submission
  JOURNAL   Submitted (15-JAN-1999) S.E. Kearsey, Oxford University, South
            Parks Road, Oxford, UK
COMMENT     On Jan 21, 1999 this sequence version replaced gi:4926.
FEATURES             Location/Qualifiers
     source          1..3667
                     /organism="Schizosaccharomyces pombe"
                     /mol_type="genomic DNA"
                     /db_xref="taxon:4896"
     gene            641..3436
                     /gene="cdc21"
     CDS             641..3436
                     /gene="cdc21"
                     /codon_start=1
                     /product="cdc21 protein"
                     /protein_id="CAA41628.1"
                     /db_xref="GI:4165293"
                     /db_xref="GOA:P29458"
                     /db_xref="UniProt/Swiss-Prot:P29458"
                     /translation="MSSSQQSGRANELRTPGRANSSSREAVDSSPLFFPASSPGSTRL
                     TTPRTTARTPLASSPLLFESSSPGPNIPQSSRSHLLSQRNDLFLDSSSQRTPRSTRRG
                     DIHSSVQMSTPSRRREVDPQRPGVSTPSSLLFSGSDALTFSQAHPSSEVADDTVRVIW
                     GTNVSIQESIASFRGFLRGFKKKYRPEYRNELMPPPDAEQLVYIEALRNMRIMGLEIL
                     NLDVQDLKHYPPTKKLYHQLYSYPQEIIPIMDQTIKDVMLDLLGTNPPEDVLNDIELK
                     IYKIRPFNLEKCINMRDLNPGDIDKLISIKGLVLRCTPVIPDMKQAFFRCSVCGHCVT
                     VEIDRGRIAEPIKCPREVCGATNAMQLIHNRSEFADKQVIKLQETPDVVPDGQTPHSV
                     SLCVYDELVDSARAGDRIEVTGIFRCVPVRLNPRMRTVKSLFKTYVDVVHIKKQDKRR
                     LGTDPSTLESDIAEDAALQIDEVRKISDEEVEKIQQVSKRDDIYDILSRSLAPSIYEM
                     DDVKKGLLLQLFGGTNKSFHKGASPRYRGDINILMCGDPSTSKSQILKYVHKIAPRGV
                     YTSGKGSSAVGLTAYITRDQDTKQLVLESGALVLSDGGICCIDEFDKMSDATRSILHE
                     VMEQQTVTVAKAGIITTLNARTSILASANPIGSKYNPDLPVTKNIDLPPTLLSRFDLV
                     YLILDRVDETLDRKLANHIVSMYMEDTPEHATDMEVFSVEFLTSYITYARNNINPVIS
                     EEAAKELVNAYVGMRKLGEDVRASEKRITATTRQLESMIRLSEAHAKMHLRNVVEVGD
                     VLEAARLIKTAIKDYATDPATGKISLDLIYVNERETLVPEDMVKELANLISNLTVGGK
                     TMLVSQLLTRFREQSSTRLDASDFEACLGAFREQSSTRLDASDFEACLGALERRGRIK
                     VITSAGHRIVRSIAQTD"
ORIGIN

Alignment Scores:
Pred. No.:              2.09e-61        Length:           3667
Score:                  1882.00         Matches:          413
Percent Similarity:     61.68%          Conservative:     168
Best Local Similarity:  43.84%          Mismatches:       242
Query Match:            40.83%          Indels:           119
DB:                     15              Gaps:             21
```

US-10-768-511-8 (1-901) x SPCDC21 (1-3667)

```
Qy     24 SerThrProLeuProGlnValThrSerProSerPheAspAsn-AlaAlaSerProValAl 43
Db    601 TCTACTCCCTTACACTGTCTTTCTACACCCACCACGAGTCATGTCCTCTAGTCAGCAAAG 660

Qy     43 aGlyArgArgAlaValArgGlnThrProThrSerAlaValArgArgArgGlyArgGluTh 63
Db    661 TGGCAGAGCCAATGAGTTGAGGACTCCT--------------------GGAAGAGCTAA 699

Qy     63 rAspSerAlaArgArgArgArgSerArgSerArgSerLeuGlyAsnSerValTyrSerSe 83
Db    700 CAGCTCTTCAAGA-----------------------------GAAGCAGTGGATTCTTC 729

Qy     83 rPro------TyrAspAlaGlyThrProGly---------------ThrProGlyThrPr 96
Db    730 TCCGTTATTTTTTCCTGCTTCGTCTCCTGGTTCTACACGCCTTACAACACCGAGGACTAC 789

Qy     96 oValAlaThrProValTyrAlaThrProValGlyThrProMetGlyThrPro-------- 113
Db    790 AGCGCGTACACCTCTCGCTTCCTCTCCTTTACTGTTTGAATCTAGTTCACCAGGTCCAAA 849

Qy    114 ----------------------------------------------------------Se 114
Db    850 CATTCCTCAGAGTTCACGTTCTCATCTTTTATCTCAGCGTAATGACTTGTTCCTTGACAG 909

Qy    114 rPheHisArgGlyThrProGlnTyrLysGlnArgSerGluLeuGlySer------GlnGl 132
Db    910 TAGTTCGCAGAGAACACCGCGCTCGACAAGACGTGGAGACATCCATTCTAGCGTGCAAAT 969

Qy    132 yLysProLeuHisArgArgArgArgSerGlnSerArgGluProGlyHisArgSerProSe 152
Db    970 GTCTACCCCATCTCGTCGAAGAGAAGTTGACCCTCAACGTCCCGGGGTAAGCACTCCTAG 1029

Qy    152 rArg------GluProSerAlaAspGlyArgProSerGluSerAlaGluPro-------- 167
Db   1030 TAGCTTGTTGTTTTCTGGGTCAGATGCACTCACGTTTAGTCAGGCGCACCCAAGCAGTGA 1089

Qy    168 -------AspAspThrLeuGlyGlyGluTyrAlaTyrValTrpGlyThrAsnValAsnIl 185
Db   1090 AGTAGCTGATGACACTGTC------------CGTGTCATTTGGGGAACAAATGTATCTAT 1137

Qy    185 eProAspValLeuArgAlaIleArgArgPheLeuHisAsnTyrArgSer----------- 201
Db   1138 TCAAGAAAGTATCGCTTCCTTTCGTGGATTTCTACGTGGGTTTAAGAAAAAGTATCGACC 1197

Qy    202 ------------------------------SerAlaHisAspLeuAsnSerLysTyrIl 211
Db   1198 TGAATACAGAAATGAACTTATGCCACCTCCTGATGCCGAACAACTA------GTGTATAT 1251

Qy    211 eGlnIleIleGluGluThrValGluArgGluGluAspThrLeuAsnIleAspMetSerAs 231
Db   1252 TGAAGCGCTACGAAACATGCGTATTATGGGTTTGGAAATATTAAATCTTGATGTTCAGGA 1311

Qy    231 pIleTyrAspHisAspPro------AspLeuTyrAlaLysIleValArgTyrProLeuAs 249
Db   1312 TTTAAAGCATTATCCCCCAACCAAGAAGCTATATCACCAGCTTTACTCATACCCTCAAGA 1371

Qy    249 pIleIleProLeuLeuAspThrGluCysGlnGluValAlaThrSerLeuLeuProThrPh 269
Db   1372 AATCATTCCTATCATGGATCAAACCATTAAAGACGTTATGCTAGACTTGTTAGGAACTAA 1431

Qy    269 eGlu-----------------------------LysHisIleGluAlaArgProPheAs 279
Db   1432 TCCTCCCGAAGATGTTCTTAATGACATTGAGCTTAAAATTTACAAAATTCGTCCTTTTAA 1491

Qy    279 nLeuLysAlaSerValHisMetArgGluLeuAsnProSerAspIleAspLysLeuValSe 299
Db   1492 TCTTGAAAAAATGCATTAATATGCGTGATCTTAATCCTGGTGACATTGACAAATTGATCTC 1551

Qy    299 rValLysGlyMetValIleArgCysSerSerIleIleProGluIleLysGlyAlaPhePh 319
Db   1552 TATTAAAGGTCTGGTACTTAGGTGTACACCTGTAATACCCGACATGAAGCAAGCATTTTT 1611

Qy    319 eLysCysLeuValCysGlyHisSerProProLeuValThrValValLysGlyArgValGl 339
Db   1612 TCGATGTTCAGTTTGTGGTCATTGT---GTTACCGTGGAAATCGATAGAGGGCGAATTGC 1668

Qy    339 uGluProThrArgCysGluLysProGluCysAlaAlaArgAsnAlaMetSerLeuIleHi 359
Db   1669 TGAGCCCATAAAATGCCCTAGAGAAGTCTGTGGTGCTACAAATGCAATGCAGCTAATTCA 1728

Qy    359 sAsnArgCysThrPheAlaAsnLysGlnIleValArgLeuGlnGluThrProAspAlaIl 379
Db   1729 TAATCGTTCTGAATTCGCCGATAAGCAAGTTATTAAACTCCAAGAGACACCTGACGTGGT 1788

Qy    379 eProGluGlyGluThrProHisThrValSerMetCysLeuTyrAsnThrMetValAspAl 399
Db   1789 ACCGGATGGCCAAACCCCACACTCTGTAAGTCTGTGCGTATATGATGAACTGGTTGACTC 1848

Qy    399 aValLysProGlyAspArgIleGluValThrGlyValPheLysAlaMetAlaValArgVa 419
Db   1849 TGCTAGAGCAGGAGATCGTATAGAAGTCACTGGGATTTTTCGCTGTGTACCAGTTCGTTT 1908

Qy    419 lGlyProAsnGlnArgThrLeuArgAlaLeuTyrLysThrTyrIleAspCysValHisVa 439
Db   1909 GAACCCTCGCATGCGTACTGTAAAATCATTATTTAAAACTTATGTCGACGTTGTTCACAT 1968

Qy    439 lLysLysSerAspArgGlyArgLeuGlnThrGluAspProMetGluMetAspLysGluAs 459
Db   1969 AAAAAAGCAGGACAAACGTCGTTTGGGTACG---GATCCTTCTACGTTG----------- 2014

Qy    459 nAspMetTyrAlaGlyTyrHisGluSerAspThrSerGluAlaAla-------------- 474
Db   2015 ----------------------GAAAGCGATATTGCTGAAGATGCTGCTTTACAAATTGA 2052

Qy    475 ----------------AsnGluAlaLysIleGlnLysLeuLysGluLeuSerLysLeuPr 489
Db   2053 TGAAGTACGGAAAATATCAGACGAAGAAGTGGAAAAAATTCAACAAGTTTCAAAGCGGGA 2112

Qy    489 oGlyIleTyrAspArgLeuSerArgSerLeuAlaProSerIleTrpGluLeuGluAspIl 509
Db   2113 CGATATATACGATATTCTTTCCCGGTCTCTGGCACCTAGCATTTACGAAATGGATGATGT 2172

Qy    509 eLysLysGlyLeuLeuCysGlnLeuPheGlyGlyLysAlaLysLysIleProSerGlyAl 529
Db   2173 TAAAAAGGGGCTTTTGCTACAGCTATTTGGTGGAACGAACAAGTCTTTTCATAAAGGCGC 2232

Qy    529 aSer------PheArgGlyAspIleAsnValLeuLeuValGlyAspProGlyThrSerLy 547
Db   2233 AAGTCCGAGATATCGTGGAGACATAAATATCTTGATGTGTGGTGATCCATCTACGTCTAA 2292

Qy    547 sSerGlnLeuLeuGlnTyrValHisLysIleAlaProArgGlyIleTyrThrSerGlyAr 567
Db   2293 GTCTCAAATTTTGAAGTATGTCCATAAGATAGCACCCAGAGGTGTTTACACTAGTGGCAA 2352

Qy    567 gGlySerSerAlaValGlyLeuThrAlaTyrValThrLysAspProGluThrArgGluTh 587
Db   2353 GGGTTCATCTGCTGTTGGTTTAACTGCCTATATTACTCGTGACCAAGACACGAAGCAACT 2412

Qy    587 rValLeuGluSerGlyAlaLeuValLeuSerAspArgGlyIleCysCysIleAspGluPh 607
Db   2413 AGTCTTGGAAAGTGGTGCTTTGGTACTGTCAGATGGGGGCATTTGCTGCATTGATGAGTT 2472

Qy    607 eAspLysMetSerAspAsnAlaArgSerMetLeuHisGluValMetGluGlnGlnThrVa 627
Db   2473 TGACAAAATGTCCGATGCTACACGTTCTATTCTTCATGAAGTTATGGAGCAGCAAACCGT 2532

Qy    627 lSerValAlaLysGlyGlyIleIleAlaSerLeuAsnAlaArgThrSerValLeuAlaCy 647
Db   2533 TACAGTTGCCAAGGCTGGCATTATTACTACTTTGAACGCTAGGACTTCTATTTTAGCTAG 2592

Qy    647 sAlaAsnProSerGlySerArgTyrAsnAlaArgLeuSerValIleAspAsnIleGlnLe 667
Db   2593 TGCAAATCCAATTGGAAGCAAATACAATCCAGACTTGCCTGTCACCAAAAATATTGATCT 2652
```

```
Qy        667 uProProThrLeuLeuSerArgPheAspLeuIleTyrLeuMetLeuAspLysProAspGl 687
              ||||||||||||||||||||||||||||||||:::|||||:::||||||:::     |||||
Db       2653 TCCGCCTACTCTCTTATCTAGATTTGATTTGGTGTATTTAATATTAGACCGCGTGGATGA 2712

Qy        687 uGlnAsnAspArgArgLeuAlaArgHisLeuValAlaLeuHisTyrGluAsn------Ty 705
              |       ||||||:::||||||     |||:::|||::::::::::   |||:::
Db       2713 AACGCTGGATAGGAAGCTCGCAAATCATATTGTTTCTATGTACATGGAAGATACTCCTGA 2772

Qy        705 rGluValSerLysGlnAspAlaLeuAspLeuGlnThrLeuThrAlaTyrIleThrTyrAl 725
                     :::       :::        ::::::   ||||||:::||||||||||||||
Db       2773 ACATGCTACAGATATGGAGGTCTTTAGTGTGGAATTTTTGACAAGCTATATCACATATGC 2832

Qy        725 aArgGlnHisValHisProThrLeuSerAspGluAlaAlaGluAspLeuIleAsnGlyTy 745
              ||||   :::::::::|||    :::|||:::|||||||||::::::|||:::|||   ||
Db       2833 TAGAAATAATATTAATCCTGTCATTAGTGAAGAGGCAGCAAAGGAACTAGTGAATGCTTA 2892

Qy        745 rValGluMetArgGlnLysGlyAsnPheProGlySerSerLysLysValIleThrAlaTh 765
              ||||   ||||||:::   |||                :::|||:::|||   |||||||||||
Db       2893 TGTTGGTATGCGAAAGCTTGGGGAGGATGTGCGAGCATCTGAGAAAAGAATTACTGCTAC 2952

Qy        765 rProArgGlnLeuGluSerMetIleArgIleSerGluAlaLeuAlaArgMetArgPheSe 785
              |   |||||||||||||||||||||||||:::|||||||||      |||:::|||
Db       2953 AACAAGACAACTGGAATCAATGATTCGTCTATCTGAGGCCCATGCGAAAATGCATCTTCG 3012

Qy        785 rGluValValGluLysValAspAlaAlaGluAlaValArgLeuLeuAspValAlaLeuGl 805
                    |||||||||      |||       ||||||   ||||||:::       |||:::::
Db       3013 CAACGTTGTTGAGGTAGGAGATGTTCTTGAAGCTGCTCGCTTGATTAAGACCGCTATTAA 3072

Qy        805 nGlnSerAlaThrAspHisAlaThrGlyThrIleAspMetAspLeuIleThrThrGlyVa 825
                 :      |||||||||   |||||||||   |||    :::|||||||||         ||
Db       3073 GGATTATGCAACTGATCCTGCAACAGGAAAGATTTCTCTTGACCTAATTTAT------GT 3126

Qy        825 lSerAlaSerGluArgIleArgArgAlaAsnLeuLeuAlaAlaLeuArgGluLeuIleAl 845
              |:::         |||    :::          :::::::::        |||     ||||||::
Db       3127 TAATGAAAGAGAGACTTTGGTACCTGAAGATATGGTCAAAGAACTGGCGAATTTAATTTC 3186

Qy        845 aAspLysIleSerProGlySerSerSerGlyLeuLysThrSerGlnLeuLeuGluAspIl 865
              ::::               |||     :::       :::       ||||||||||||
Db       3187 AAATTTAACTGTGGGTGGAAAAACT------ATGTTGGTCAGTCAATTGTTAACTCGGTT 3240

Qy        865 eArgSerGlnSerSerValAspValSerLeuGlnAspIleLysAsnAlaLeuGlySerLe 885
                 |||   |||||||||        :::            |||   :::       ||||||:::
Db       3241 TCGTGAGCAGAGTTCTACTCGCTTAGATGCTTCTGATTTCGAAGCTTGTCTGGGTGCTTT 3300

Qy        885 uGln 886
                :::
Db       3301 TCGT 3304
```

```
RESULT 9
AC149637
LOCUS       AC149637              116616 bp    DNA     linear   PLN 20-MAY-2005
DEFINITION  Medicago truncatula clone mth2-181h2, complete sequence.
ACCESSION   AC149637
VERSION     AC149637.11  GI:66351814
KEYWORDS    HTG.
SOURCE      Medicago truncatula (barrel medic)
  ORGANISM  Medicago truncatula
            Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
            Spermatophyta; Magnoliophyta; eudicotyledons; core eudicotyledons;
            rosids; eurosids I; Fabales; Fabaceae; Papilionoideae; Trifolieae;
            Medicago.
REFERENCE   1  (bases 1 to 116616)
  AUTHORS   Lin,S., Dixon,R., May,G., Sumner,L., Gonzales,B., Cook,D., Kim,D.
            and Roe,B.A.
  TITLE     Medicago truncatula BAC Clone mth2-181h2
  JOURNAL   Unpublished
REFERENCE   2  (bases 1 to 116616)
  AUTHORS   Lin,S., Dixon,R., May,G., Sumner,L., Gonzales,B., Cook,D., Kim,D.
            and Roe,B.A.
  TITLE     Direct Submission
  JOURNAL   Submitted (17-JUN-2004) Department Of Chemistry And Biochemistry,
            The University Of Oklahoma, 620 Parrington Oval, Room 208, Norman,
            OK 73019, USA
REFERENCE   3  (bases 1 to 116616)
  AUTHORS   Lin,S., Dixon,R., May,G., Sumner,L., Gonzales,B., Cook,D., Kim,D.
            and Roe,B.A.
  TITLE     Direct Submission
  JOURNAL   Submitted (30-APR-2005) Department Of Chemistry And Biochemistry,
            The University Of Oklahoma, 620 Parrington Oval, Room 208, Norman,
            OK 73019, USA
REFERENCE   4  (bases 1 to 116616)
  AUTHORS   Lin,S., Dixon,R., May,G., Sumner,L., Gonzales,B., Cook,D., Kim,D.
            and Roe,B.A.
  TITLE     Direct Submission
  JOURNAL   Submitted (20-MAY-2005) Department Of Chemistry And Biochemistry,
            The University Of Oklahoma, 620 Parrington Oval, Room 208, Norman,
            OK 73019, USA
COMMENT     On May 20, 2005 this sequence version replaced gi:62988494.
            -------------- Genome Center
            Center: Department Of Chemistry And Biochemistry
            The University Of Oklahoma
            Center code:UOKNOR
            ---------------------.
FEATURES             Location/Qualifiers
     source          1..116616
                     /organism="Medicago truncatula"
                     /mol_type="genomic DNA"
                     /db_xref="taxon:3880"
                     /clone="mth2-181h2"
                     /clone_lib="Medicago truncatula BAC library H2"
ORIGIN

Alignment Scores:
Pred. No.:              4.8e-59        Length:         116616
Score:                  1858.50        Matches:        505
Percent Similarity:     37.18%         Conservative:   140
Best Local Similarity:  29.11%         Mismatches:     182
Query Match:            40.32%         Indels:         912
DB:                     15             Gaps:           26

US-10-768-511-8 (1-901) x AC149637 (1-116616)

Qy         12 ValSerSerProTyrProSerGlnSer------------------------------- 20
              ||||||||||||:::      :::
Db      42167 GTTTCATCTCCACATTCTTATTCTAATTTCTAATTTCTAATTTCAATTTTCTGAAACCCT 42226

Qy         21 ------------GluGlyValSerThrProLeuProGlnValThrSerProSerPheAsp 36
                          |||   |||:::|||          :::||||||      :::
Db      42227 ATTTTTCACTTTTCAGGTCCTTCTTCACCGGACAACTCCAATTCAAGTCCAATCGCCAAC 42286

Qy         37 AsnAlaAlaSerProValAlaGlyArgArgAlaValArgGlnThrProThrSerAlaVal 56
                 :::||||||      |||||||||
Db      42287 ACCTACTCCTCCCCC------GGCCGCCGT------------------------------ 42310

Qy         57 ArgArgArgGlyArgGluThrAspSerAlaArgArgArgArgSerArgSerArgSerLeu 76
              |||||||||          |||   |||         :::   :::|||||||||
Db      42311 CGCCGCCGTCAATCAATCACTCCATCTGGAACTCCTCAACCGAACAGATCTCGC------ 42364

Qy         77 GlyAsnSerValTyrSerSerProTyrAspAlaGlyThrProGlyThrProGlyThrPro 96

Db      42364 ------------------------------------------------------------ 42364

Qy         97 ValAlaThrProValTyrAlaThrProValGlyThrProMetGlyThrProSerPheHis 116
                       :::|||:::          ||||||      |||||||||   |||
Db      42365 ---------------TTCGCTTCCTCCGATGCAACTCCA------ACTCCCTCCCGTCAC 42403

Qy        117 ArgGlyThrProGlnTyrLysGlnArgSerGluLeuGlySerGlnGlyLysProLeuHis 136
              |||                                       |||   |||
Db      42404 CGT--------------------------------CGAGGTGGCGGAATCCCCTCCGGC 42430

Qy        137 ArgArgArgArgSerGlnSerArgGluProGlyHisArgSerProSerArgGluProSer 156
              ||||||         :::                              |||
Db      42431 CGCCGTGGACCCGCCACACCAACTTCCACCTCCGAAGACTTTCCTATGTCTTCTGAAGGC 42490

Qy        157 AlaAspGlyArgProSerGluSerAlaGluProAspAspThrLeuGlyGlyGluTyrAla 176
                 ||||||            :::   |||   |||
```

```
Db      42491 GGTGACGGTTATGACATGGATGATGCTGGTCCT------------------------ACT 42526

Qy        177 TyrValTrpGlyThrAsnValAsnIleProAspValLeuArgAlaIleArgArgPheLeu 196

Db      42527 TACGTTTGGGGTACTAACATTAGCGTGGAGGATGTTAACGATGCGATTCAGAGGTTTTTG 42586

Qy        197 HisAsnTyrArgSerSerAlaHis------------------------------------ 204

Db      42587 AAGCATTTTAGAGAAAAATCTGCTTCTCAAGGCGGTGATGATGATTTGGATATGGATTTG 42646

Qy        205 AspLeuAsnSerLysTyrIleGlnIleIleGluGluThrValGluArgGluGluAspThr 224

Db      42647 GATATTGAAGGGAAGTATGAGAAATTGATCAAACAGGTTATTGAATTGGAAGGTGAGTCT 42706

Qy        225 LeuAsnIleAspMetSerAspIleTyrAspHisAspProAspLeuTyrAlaLysIleVal 244

Db      42707 ATTGATGTTGATGCACGTGATGTGTTTGATCACGATCACGAATTGTATAGTAAGATGGTG 42766

Qy        245 ArgTyrProLeuAspIleIleProLeuLeuAspThrGluCysGlnGluValAlaThrSer 264

Db      42767 AGGTATCCTCTCGAGGTGCTTGCCATTTTTGATATGGTTTTGATGAACATGGTCACTAGG 42826

Qy        265 LeuLeuProThrPheGluLysHisIleGluAlaArgProPheAsnLeuLysAlaSerVal 284

Db      42827 ATGAATCCTATGTTTGAGAAGCATGTTCAAACTAGGATTTTAATCTTAAGAGCTCAACT 42886

Qy        285 HisMetArgGluLeuAsnProSer------------------------------------ 292

Db      42887 TCAATGAGGAATTTGAATCCTTCTGGTGAGTTTTTGTTTTTTGAAATTATTGTTATTGA 42946

Qy        292 ------------------------------------------------------------ 292

Db      42947 ATTCAATTTTGTGTGTTTGATTTTAGCAATTAGATTGATATGTAATGATTGTTTCTTGAT 43006

Qy        293 ------------AspIleAspLysLeuValSerValLysGlyMetValIleArgCysSer 308

Db      43007 GTTTGTTTTGAAGATATTGAGAGGATGGTGTCGATGAAGGGAATGATTATTCGGTGCAGC 43066

Qy        309 SerIleIleProGluIleLysGlyAlaPhePheLysCysLeuValCysGlyHisSerPro 328

Db      43067 TCGATTATACCAGAAATCAGGGAAGCTATATTCAGGTGTCTAGTATGTGGTTTCTGCTCT 43126

Qy        329 ProLeuValThrValValLys--------------------------------------- 335

Db      43127 GACCCGGTTCCGGTGGAGAGAGGTAATTACAGAGTCATTTCTGGTTATTTATGTAGTTAA 43186

Qy        336 ---------------------------------------------------GlyArgVal 338

Db      43187 CTGTTGTGTTTATATCTGAATGTAATATTTTGCTGGCTATTTTGATTTTAGGACGAATC 43246

Qy        339 GluGluProThrArgCysGluLysProGluCysAlaAlaArgAsnAlaMetSerLeuIle 358

Db      43247 GCCGAACCTACTGTTTGCTTGAGGGAAGAGTGTCAGTCCAGGAATTCGATGACACTTGTT 43306

Qy        359 HisAsnArgCysThr--------------------------------------------- 363

Db      43307 CACAACCGATGCAA-GTAAGTCTGGCTTATTTGTTTAAAGAAGTGATTTTTTTAGTTGA 43365

Qy        363 ------------------------------------------------------------ 363

Db      43366 TTATTTGGAATTGGAGTACTATATTTTGACTGTTGGATAGTGTTGTAACTTGTATGCATC 43425

Qy        363 ------------------------------------------------------------ 363

Db      43426 ATATTTGATATTGAAGGGGCTATATTCTGAAGCAAGCTTCTTTTAACCTTTTTCCTTTAA 43485

Qy        363 ------------------------------------------------------------ 363

Db      43486 CTCACATGCTATGTATATGATGATTAGTCAGTTCCTTGTTCTTTCTGATTTTTGGTTGTG 43545

Qy        363 ------------------------------------------------------------ 363

Db      43546 AGCAATCAGTGGCAGCACTGAATTTTCCTGCATAGTAAAAGATAGGCAAATAGCTGGATA 43605

Qy        363 ------------------------------------------------------------ 363

Db      43606 TGGAAGTTAACTTTCTGAACTGGTCTACTACTAAATTATTTCCATGCAACAACGACTAAT 43665

Qy        363 ------------------------------------------------------------ 363

Db      43666 AACCTGCTGGTGGTTTAGTGTAAAATTGACCAATGTCTGGCTAAAATTGGGTTAATTTTT 43725

Qy        363 ------------------------------------------------------------ 363

Db      43726 TACAAAAAATCAATCTGATCTAGAAGGTGAGGAACTGTGCTTTATTTGTCAAGAGGTTAC 43785

Qy        363 ------------------------------------------------------------ 363

Db      43786 GATAAAAACAGTTTGAAAAATGATTATTTGAACTGTACTGGGTATATTTTAGTGATATGT 43845

Qy        363 ------------------------------------------------------------ 363

Db      43846 TGGATGCTTGCCATGGTAATTATCCCCTGTTTTAAACAATGTGTTTTGGTCATGTGACTA 43905

Qy        363 ------------------------------------------------------------ 363

Db      43906 TTTGGCGTAAGGTTTTTCACTTGGATTTCAGTTCATAAACCTTGAAACTGCAATTAAATT 43965

Qy        364 ------------------PheAlaAsnLysGlnIleValArgLeuGlnGluThrProAsp 377

Db      43966 ATCTTCTCTAATTTCAGGTTTTCTGATAAACAAATTGTGAGGCTTCAGGAGACACCTGAT 44025

Qy        378 AlaIleProGluGlyGluThrProHisThrValSerMetCysLeuTyrAsnThrMetVal 397

Db      44026 GAGATACCTGAAGGAGGAACACCTCACACTGTTAGCTTGCTATTGCATGACAAACTGGTG 44085

Qy        398 AspAlaValLysProGlyAspArgIleGlu------------------------------ 407

Db      44086 GATACTGGAAAGCCAGGTGACAGAGTTGAGGTGAGACTATACCATGAGATACTCTGATAA 44145

Qy        408 ------------------------------------------------------ValThr 409

Db      44146 ATAGATATTCTTATGTTTGTTCAATCGTCCCATGAGTGTTTTGTTGTACATCAGGTGACT 44205

Qy        410 GlyValPheLysAlaMetAlaValArgValGlyProAsnGlnArgThrLeuArgAlaLeu 429

Db      44206 GGTATATACAGGGCTATGAGTGTCAGGGTGGGACCAACTCAGAGAACTGTTAAGTCATTG 44265

Qy        430 TyrLys------------------------------------------------------ 431

Db      44266 TTCAAAGTATGGGTTACCTTCTCAAACTCATTTGCATTTAGAGCTTCTCCTTTTCATTTT 44325

Qy        432 ---------------------------------------------ThrTyrIleAspCys 436

Db      44326 TCATCTTGTCTAATTGACATTTGTTTTCTTTTGTTCATGTTGTAGACGTATATTGATTGT 44385

Qy        437 ValHisValLysLysSerAspArgGlyArgLeuGlnThrGluAspProMetGluMetAsp 456

Db      44386 CTTCACATAAAAAAGACATCAAAGTCCCGGATGCTAGTTGAGGATGCTATGGAGGTTGAC 44445

Qy        457 LysGlu------------------------------------------------------ 458

Db      44446 AATGGCCAGGGCAGAAATCCAGAAGAAGTTTTATTTGATGAAGAAAAGGTACATTTTGTA 44505

Qy        459 ---------AsnAspMetTyrAlaGlyTyrHisGluSerAspThrSerGluAlaAlaAsn 475

Db      44506 CATCTCTTCAATGAATTATCTGCTTGTTTATTCACCCGCTAACTGCA---TGCTTGTCT 44562

Qy        476 GluAlaLysIleGlnLysLeuLysGluLeuSerLysLeuProGlyIleTyrAspArgLeu 495

Db      44563 GAGTGTCAGGTGGCTAAACTGAAGGAGCTCTCAAAACAACCAGATATCTATGAAAGGCTG 44622

Qy        496 SerArgSerLeuAlaProSerIleTrpGluLeuGluAspIleLysLysGlyLeuLeuCys 515

Db      44623 ACAAAGTCACTGGCGCCAAACATCTGGGAGCTAGATGATGTAAAGAAAGGTCTTCTTTGC 44682
```

```
Qy        515 ------------------------------------------------------------ 515
Db      44683 CAGGTAAGACATCTTTCTAGCTAGTCTTTCATTTGTTGTTTTACAAATATTCTGAGGATG 44742
Qy        516 ------------------GlnLeuPheGlyGlyLysAlaLysLysIleProSerGlyAla 529
Db      44743 CTGGTTTTTATTTTCTTGCAGCTTTTGGGGGCAATGCTTTGAAGTTGGCATCTGGTGCT 44802
Qy        530 SerPheArgGlyAspIleAsnValLeuLeuValGlyAspProGlyThrSerLysSerGln 549
Db      44803 AGCTTTCGTGGTGATATAAATATTCTCCTTGTTGGTGATCCTGGTACAAGCAAGTCGCAG 44862
Qy        550 LeuLeuGlnTyrValHisLysIleAlaProArgGlyIleTyrThrSerGlyArgGlySer 569
Db      44863 CTACTTCAGTACATACACAAACTTTCTCCACGTGGCATTTACACCAGTGGAAGGGGAAGC 44922
Qy        570 SerAlaValGlyLeuThrAlaTyrValThrLysAspProGluThrArgGluThr------ 587
Db      44923 TCAGCAGTTGGATTGACTGCTTACGTGGCCAAAGATCCTGAAACAGGCGAAACTGTATAG 44982
Qy        587 ------------------------------------------------------------ 587
Db      44983 ACCAAGTCCATGATCTTTAGTGTCTATAGTACAAGAAATGCAATGGAGTCTAATTTTTT 45042
Qy        588 ---------------ValLeuGluSerGlyAlaLeuValLeuSerAspArgGlyIleCys 602
Db      45043 CTGCTAAATCTTCAGGTTCTTGAGAGTGGTGCCCTAGTTTTGAGCGACCGAGGTATCTGC 45102
Qy        603 CysIleAspGluPheAspLysMetSerAspAsnAlaArgSerMetLeuHisGlu------ 620
Db      45103 TGCATTGATGAATTTGACAAAATGTCTGACAATGCAAGGAGCATGTTACACGAGGTTGGT 45162
Qy        620 ------------------------------------------------------------ 620
Db      45163 TTCATTTTTTGTCAGCATGTATTTCTGCAGTCCTATCAGCTTGTTTTTTCTACAGAACTT 45222
Qy        620 ------------------------------------------------------------ 620
Db      45223 TGCATGGATTGAAAAATGGGTAATGTTGGCACTAGTTAAGGAGACTAATTTAGAAATTTT 45282
Qy        620 ------------------------------------------------------------ 620
Db      45283 GCATTGGAAATTGTTAAAAAGTGGTATGTTCAACTTTTTAAAAGTCAAAAGTTTGTTGTT 45342
Qy        620 ------------------------------------------------------------ 620
Db      45343 TTCAAGTAAAAATTTCTATTTTTTACTTTCTTAACTAGTGCCCCGGGGGCACTAGTTAAC 45402
Qy        620 ------------------------------------------------------------ 620
Db      45403 ATGACCCTTGAAAAATATATACACTATGATTTCGTGCTTTATCATGTTAGCTTTGTACTT 45462
Qy        621 ------------------------------------ValMetGluGlnGlnThrValSer 628
Db      45463 TTGAGTAACTAATACACTTGTCTCTGTTTATTCCAGGTGATGGAACAACAAACTGTTTCA 45522
Qy        629 ValAlaLysGlyGlyIleIleAlaSerLeuAsnAlaArgThrSerValLeuAlaCysAla 648
Db      45523 ATAGCAAAGGCAGGCATTATTGCTTCCCTCAATGCAAGGACCTCGGTGTTGGCTTGCGCA 45582
Qy        649 AsnProSerGlySerArgTyrAsnAlaArgLeuSerValIleAspAsnIleGlnLeuPro 668
Db      45583 AACCCTAGTGGGTCACGATATAATCCCCGCTTGTCTGTCATTGACAACATACACCTTCCT 45642
Qy        669 ProThrLeuLeuSer--------------------------------------------- 673
Db      45643 CCCACTCTTTTGTCCAGGTTAGTTGTTGGACATAAGCCGTGATAACTAGGTTATGGATCT 45702
Qy        673 ------------------------------------------------------------ 673
Db      45703 GGCTACGTATTTATGCAATTCCTCATGCATTGGCAGCAATCAATTTTATACCTTGGGCCT 45762
Qy        673 ------------------------------------------------------------ 673

Db      45763 TACCTCTATTTTTTTGGGTGTTTATTTGAACATAACACAAAATTGAATCTGTTGTTGATT 45822
Qy        673 ------------------------------------------------------------ 673
Db      45823 CTCTTTATCAAACCTAGTGTGGGTAGGGGTTGGCATAAAGCAAAAAATTCTCTTTAAAAA 45882
Qy        673 ------------------------------------------------------------ 673
Db      45883 TAATGGAAGCTACTAGACAGTTTTGCAAGTGTTGCTATTGATTTAAGTTGATCTTTTTCC 45942
Qy        674 ----ArgPheAspLeuIleTyrLeuMetLeuAspLysProAspGluGlnAsnAspArgAr 692
Db      45943 TGGCAGATTTGATTTAATTTACTTAATGCTTGATAAGGCTGATGAACAGACCGACCGACG 46002
Qy        692 gLeuAlaArgHisLeuValAlaLeuHisTyrGluAsnTyrGluValSer----------- 708
Db      46003 ACTTGCCAAACATATTGTTTCCCTGCACTATAAGGATTATGAGGTCTG-TTTATAATCTT 46061
Qy        708 ------------------------------------------------------------ 708
Db      46062 GTAATCATCTTATTCAAGATTAAGATCCAAACTTAAGTGTTTTTGCATACTCATTTAGAA 46121
Qy        709 ---------------------------LysGlnAspAlaLeuAspLeuGlnThrLeuTh 719
Db      46122 TGATATGTGGCCATAATTGCAGAACATTGAGCAGGATGTGTTGGACATTTCTACACTAAC 46181
Qy        719 rAlaTyrIleThrTyrAlaArgGlnHisValHisProThrLeuSerAspGluAlaAlaGl 739
Db      46182 TGATTATGTTAGCTATGCCCGAAAGCACATACACCCACAACTATCAGATGAAGCTGCTGA 46241
Qy        739 uAspLeuIleAsnGlyTyrValGluMetArgGlnLysGlyAsnPheProGlySerSerLy 759
Db      46242 TGAGTTGATTACAGGCTATGTAAAAATAAGGGGTAGGGGAAAATTCTCCGGAAGCAGCAA 46301
Qy        759 sLysVal----------------------------------------------------- 761
Db      46302 AAAGGT-GATACAAAGTGAATTACTTGATCTTAGAATTATGCTTATTTTCATTTTTTTTC 46360
Qy        762 ----------------------------------------------------IleTh 763
Db      46361 ATGTTTATCTGTTGTGTATGTTAATCTTGCCTATCTATCAATTAATTAACAGGTAATAAC 46420
Qy        763 rAlaThrProArgGlnLeuGluSerMetIleArgIleSerGluAlaLeuAlaArgMetAr 783
Db      46421 GGCAACTCCAAGGCAGATTGAGAGTCTATTACGACTTAGTGAAGCATTGGCTCGGATTCG 46480
Qy        783 gPheSerGluVal----------------------------------------------- 787
Db      46481 CTTCTCAGAATC-GGTTAGTTGAGCTTGCTTGGTCTCTTCCTGATTTGTAGATAAAAAAT 46539
Qy        788 ---------------------------------ValGluLysValAspAlaAlaGluAl 796
Db      46540 AGCATCCAGAATTATTAAGTAACTTGTTGCACAGGTCGAAAAACATGACGTGGTGGAGGC 46599
Qy        796 aValArgLeuLeuAspValAlaLeuGlnGlnSerAlaThrAsp----------------- 810
Db      46600 ATTTCGGCTTCTGGAAGTTGCAATGCAGCAATCTGCAATGGATACCAGAACTGGTAGTTA 46659
Qy        811 ------------------------------HisAla------------------------ 812
Db      46660 TCCAAACAAGCCCCTATACTCTATTGTATTCCACCAAATAAAACTATATATCAGACTGAA 46719
Qy        813 -------------ThrGlyThrIleAspMetAspLeuIleThrThrGlyValSerAlaSe 828
Db      46720 TCCTTTTATCTTCACAGGAACTATTGACATGGATCTTATAACTACTGGTGTTTCGGCAAG 46779
Qy        828 rGluArgIleArgArgAlaAsnLeuLeuAlaAlaLeuArgGluLeuIleAlaAspLysIl 848
Db      46780 CGAAAGGATGAGGCGAGAAAGTTTGATTCAAGACACCCGCAACATGATCATGGAGAAGAT 46839
Qy        848 eSerProGlySerSerSerGlyLeuLysThrSerGlnLeuLeuGluAspIleArgSerGl 868
```

```
Db      46840 GCAAATTGGGGGACGTTCC---ATGCGATTATTAGAGGTATTACCATTCTTTCTAAAACC 46896

Qy        868 nSerSerValAspVal------------------------------------------- 873
                    |||:::|||
Db      46897 ATTCAGTATAGATTATGTTCAGTTTGAGACAATGATCTAATATAATTTAACACGATCACG 46956

Qy        873 ------------------------------------------------------------ 873

Db      46957 CAGATATTGGAGGAGTTGAAGAATCAGAGCCCTGGCAGCGAAATTCATCTTAATGATGTA 47016

Qy        873 ------------------------------------------------------------ 873

Db      47017 AGCCCTTTCCTTTGAAATTATTTACCAGTGTATATGCTTTTGCTCCATCCCCGAGGATA 47076

Qy        874 ---------------------SerLeuGlnAspIleLysAsnAlaLeuGlySerLeuGl 886
                                   |||     :::::::::|||:::   :::|||
Db      47077 AGTGAACTAATGAGCTATTTCATGTTTAATGCAGCTAAGGAGTGCAGTTGCTACTCTTGC 47136

Qy        886 nGlyGluGlyPheLeuThrValHisGlyAspIleValLysArg 900
                 ||||||||:::::::::   ||||||    |||||||||
Db      47137 CAGTGAAGGATTTGTTTCCATGGCTGGTGACAGTGTAAAAAGA 47179
```

RESULT 10
SCU14731

```
LOCUS       SCU14731                5392 bp    DNA     linear   PLN 31-JAN-1996
DEFINITION  Saccharomyces cerevisiae Cdc54p (CDC54) gene, complete cds.
ACCESSION   U14731
VERSION     U14731.1  GI:608170
KEYWORDS    .
SOURCE      Saccharomyces cerevisiae (baker's yeast)
  ORGANISM  Saccharomyces cerevisiae
            Eukaryota; Fungi; Ascomycota; Saccharomycotina; Saccharomycetes;
            Saccharomycetales; Saccharomycetaceae; Saccharomyces.
REFERENCE   1  (bases 1 to 5392)
  AUTHORS   Whitebread,L.A. and Dalton,S.
  TITLE     Cdc54 belongs to the Cdc46/Mcm3 family of proteins which are
            essential for initiation of eukaryotic DNA replication
  JOURNAL   Gene 155 (1), 113-117 (1995)
   PUBMED   7698653
REFERENCE   2  (bases 1 to 5392)
  AUTHORS   Dalton,S.
  TITLE     Direct Submission
  JOURNAL   Submitted (15-SEP-1994) Stephen Dalton, Department of Gene
            Regulation, Roche Institute of Molecular Biology, 340 Kingsland
            Street, Nutley, NJ 07110-1199, USA
FEATURES             Location/Qualifiers
     source          1..5392
                     /organism="Saccharomyces cerevisiae"
                     /mol_type="genomic DNA"
                     /db_xref="taxon:4932"
     gene            1921..4722
                     /gene="CDC54"
     CDS             1921..4722
                     /gene="CDC54"
                     /codon_start=1
                     /product="Cdc54p"
                     /protein_id="AAA86310.1"
                     /db_xref="GI:608171"
                     /translation="MSQQSSSPTKEDNNSSSPVVPNPDSVPPQLSSPALFYSSSSSQG
                     DIYGRNNSQNLSQGEGNIRAAIGSSPLNFPSSSQRQNSDVFQSQGRQGRIRSSASASG
                     RSRYHSDLRSDRALPTSSSSLGRNGQNRVHMRRNDIHTSDLSSPRRIVDFDTRSGVNT
                     LDTSSSSAPPSEASEPLRIIWGTNVSIQECTTNFRNFLMSFKYKFRKILDEREEFINN
                     TTDEELYYIKQLNEMRELGTSNLNLDARNLLAYKQTEDLYHQLLNYPQEVISIMDQTI
                     KDCMVSLIVDNNLDYDLDEIETKFYKVRPYNVGSCKGMRELNPNDIDKLINLKGLVLR
                     STPVIPDMKVAFFKCNVCDHTMAVEIDRGVIQEPARCERIDCNEPNSMSLIHNRCSFA
                     DKQVIKLQETPDFVPDGQTPHSISLCVYDELVDSCRAGDRIEVTGTFRSIPIRANSRQ
                     RVLKSLYKTYVDVVHVKKVSDKRLDVDTSTIEQELMQNKVDHNEVEEVRQITDQDLAK
                     IREVAAREDLYSLLARSIAPSIYELEDVKKGILLQLFGGTNKTFTKGGRYRGDINILL
                     CGDPSTSKSQILQYVHKITPRGVYTSGKGSSAVGLTAYITRDVDTKQLVLESGALVLS
                     DGGVCCIDEFDKMSDSTRSVLHEVMEQQTISIAKAGIITTLNARSSILASANPIGSRY
                     NPNLPVTENIDLPPPLLSRFDLVYLVLDKVDEKNDRELAKHLTNLYLEDKPEHISQDD
                     VLPVEFLTMYISYAKEHIHPIITEAAKTELVRAYVGMRKMGDDSRSDEKRITATTRQL
                     ESMIRLAEAHAKMKLKNVVELEDVQEAVRLIRSAIKDYATDPKTGKIDMNLVQTGKSV
                     IQRKLQEDLSREIMNVLKDQASDSMSFNELIKQINEHSQDRVESSDIQEALSRLQQED
                     KVIVLGEGVRRSVRLNNRV"
ORIGIN

Alignment Scores:
Pred. No.:              2.03e-59      Length:         5392
Score:                  1833.00       Matches:        393
Percent Similarity:     59.36%        Conservative:   178
Best Local Similarity:  40.85%        Mismatches:     281
Query Match:            39.77%        Indels:         110
DB:                     15            Gaps:           14

US-10-768-511-8 (1-901) x SCU14731 (1-5392)

Qy          2 GluAsnAsnAspAlaLeuAspIleGlyAlaValSerSerProTyrProSerGlnSerGlu 21
                |||:::|||::::::                         |||||||||           :::
Db       1951 GAGGATAATAACTCC------------------AGCTCCCCCGTTGTGCCTAATCCTGAT 1992

Qy         22 GlyValSerThrProLeuProGlnValThrSerProSerPheAspAsnAlaAlaSerPro 41
                 |||               ||||||::::::|||||||:::          ::::::|||
Db       1993 TCTGTTCCA---------CCACAGCTTTCTTCCCCAGCTCTATTTTATAGCTCCTCTTCA 2043

Qy         42 ValAlaGly--------------------------------------------------- 44
                    |||
Db       2044 TCACAAGGTGATATCTATGGTCGCAACAATAGCCAGAACTTAAGTCAGGGAGAGGGAAAC 2103

Qy         45 ---ArgArgAlaValArgGlnThrPro---------ThrSerAlaValArgArgArgGly 60
                   |||   |||:::      :::|||         |||:::   :::|||:::
Db       2104 ATCAGAGCTGCTATAGGTTCTTCTCCACTAAATTTTCCATCTTCTTCCCAAAGACAAAAT 2163

Qy         61 ArgGluThrAspSerAlaArgArgArgArgSerArgSerArgSerLeuGlyAsnSerVal 80
                 :::      ::::::   |||:::   |||   ||||||          ::::::
Db       2164 TCCGATGTTTTCCAATCTCAAGGCAGACAGGGCAGAATTCGTTCTTCTGCGAGCGCTTCT 2223

Qy         81 TyrSerSerProTyrAspAlaGlyThrProGlyThrProGlyThrProValAlaThrPro 100
                    |||   |||   :::                          |||   ::::::
Db       2224 GGAAGGTCTAGATATCATTCTGATCTGAGAAGTGATAGAGCACTGCCTACTTCTTCC--- 2280

Qy        101 ValTyrAlaThrProValGlyThrProMetGlyThrProSerPheHisArgGlyThrPro 120

Db       2280 ------------------------------------------------------------ 2280

Qy        121 GlnTyrLysGlnArgSerGluLeuGlySerGlnGlyLysProLeuHisArgArgArgArg 140
                       |||   ||||||      |||:::                     ||||||
Db       2281 ---------------TCTTCTTTAGGCCGGAATGGTCAAAACCGTGTACACATGCGAAGA 2325

Qy        141 SerGlnSerArgGluProGlyHisArgSerProSerArgGluProSerAlaAspGlyArg 160
              :::                        ||||||   |||               |||   |||
Db       2326 AATGATATTCATACATCTGATTTATCCTCTCCTAGAAGAATTGTGGATTTTGATACTAGA 2385

Qy        161 ------------------------------------------ProSerGluSerAlaGlu 166
                                                         |||||||||::::::|||
Db       2386 TCCGGCGTGAATACTTTGGATACTTCTTCTTCTTCTGCTCCTCCATCGGAAGCCAGTGAA 2445

Qy        167 ProAspAspThrLeuGlyGlyGluTyrAlaTyrValTrpGlyThrAsnValAsnIlePro 186
              |||                         :::||||||||||||||||||:::|||
Db       2446 CCC---------------------TTGAGAATAATTTGGGGTACCAACGTCAGTATCCAG 2484

Qy        187 AspValLeuArgAlaIleArgArgPheLeuHisAsnTyrArg------------------ 200
              :::                  |||   ||||||   :::::::::
Db       2485 GAATGTACAACTAATTTTCGTAATTTTTTGATGTCTTTCAAGTATAAATTCCGTAAAATA 2544

Qy        201 ------------------------SerSerAlaHisAspLeuAsnSerLysTyrIleGln 212
                                      ::::::      |||            ||||||:::
Db       2545 TTGGATGAAAGGGAGGAATTCATTAACAATACTACCGACGAAGAACTATACTATATCAAG 2604

Qy        213 IleIleGluGluThrValGluArgGluGluAspThrLeuAsnIleAspMetSerAspIle 232
                 :::   |||      |||               |||||||:::|||      ::::::
Db       2605 CAGCTTAATGAAATGAGAGAACTTGGTACCTCTAATTTAAACTTGGATGCTAGAAACTTA 2664

Qy        233 ------TyrAspHisAspProAspLeuTyrAlaLysIleValArgTyrProLeuAspIle 250
                    |||         |||||||||   :::::::::   ||||||   ::::::
Db       2665 CTTGCTTACAAGCAAACAGAAGACTTATACCATCAACTATTAAATTATCCTCAAGAAGTG 2724
```

```
Qy    251 IleProLeuLeuAspThrGluCysGlnGluValAlaThrSerLeu--------------- 265
Db   2725 ATTTCTATCATGGATCAAACTATCAAGGACTGTATGGTTTCCTTAATAGTAGACAACAAT 2784

Qy    266 ------------LeuProThrPheGlu---LysHisIleGluAlaArgProPheAsnLeu 280
Db   2785 TTGGATTACGATTTAGATGAAATAGAGACCAAGTTTTATAAAGTAAGGCCTTACAATGTA 2844

Qy    281 LysAlaSerValHisMetArgGluLeuAsnProSerAspIleAspLysLeuValSerVal 300
Db   2845 GGTTCTTGTAAAGGTATGCGCGAATTGAATCCAAATGATATTGATAAATTGATAAACTTA 2904

Qy    301 LysGlyMetValIleArgCysSerSerIleIleProGluIleLysGlyAlaPhePheLys 320
Db   2905 AAGGGTCTTGTCCTGAGATCAACTCCGGTTATCCCTGATATGAAGGTAGCGTTTTTCAAA 2964

Qy    321 CysLeuValCysGlyHisSerProProLeuValThrValValLysGlyArgValGluGlu 340
Db   2965 TGCAACGTCTGCGATCATACA---ATGGCAGTGGAAATTGATAGAGGAGTTATACAAGAG 3021

Qy    341 ProThrArgCysGluLysProGluCysAlaAlaArgAsnAlaMetSerLeuIleHisAsn 360
Db   3022 CCCGCTAGGTGTGAACGTATTGATTGTAATGAACCCAATTCCATGTCACTGATTCACAAC 3081

Qy    361 ArgCysThrPheAlaAsnLysGlnIleValArgLeuGlnGluThrProAspAlaIlePro 380
Db   3082 AGGTGTTCATTCGCAGATAAACAAGTCATTAAGTTACAGGAAACTCCAGACTTTGTGCCT 3141

Qy    381 GluGlyGluThrProHisThrValSerMetCysLeuTyrAsnThrMetValAspAlaVal 400
Db   3142 GATGGACAAACGCCTCACTCTATCTCATTATGTGTTTACGATGAATTAGTGGATTCTTGC 3201

Qy    401 LysProGlyAspArgIleGluValThrGlyValPheLysAlaMetAlaValArgValGly 420
Db   3202 AGGGCGGGCGATCGTATTGAGGTGACTGGCACGTTCAGGTCCATCCCCATTAGAGCTAAT 3261

Qy    421 ProAsnGlnArgThrLeuArgAlaLeuTyrLysThrTyrIleAspCysValHisValLys 440
Db   3262 TCCAGGCAACGCGTACTAAAGTCGTTGTATAAAACATACGTCGATGTGGTCCACGTTAAA 3321

Qy    441 LysSerAspArgGlyArgLeuGlnThrGluAspProMetGluMetAspLysGluAsnAsp 460
Db   3322 AAAGTTTCAGATAAGAGGTTAGACGTCGATACTTCTACTATTGAACAAGAATTAATGCAG 3381

Qy    461 MetTyrAlaGlyTyrHisGluSerAspThrSerGluAlaAlaAsnGluAlaLysIleGln 480
Db   3382 AACAAGGTAGATCATAACGAGGTCGAAGAAGTAAGACAAATTACTGATCAGGATTTAGCA 3441

Qy    481 LysLeuLysGluLeuSerLysLeuProGlyIleTyrAspArgLeuSerArgSerLeuAla 500
Db   3442 AAAATTCGCGAGGTTGCGGCAAGAGAAGATTTATACAGTTTATTAGCCCGCTCTATTGCC 3501

Qy    501 ProSerIleTrpGluLeuGluAspIleLysLysGlyLeuLeuCysGlnLeuPheGlyGly 520
Db   3502 CCAAGTATTTACGAGCTAGAAGATGTCAAGAAGGGTATATTACTTCAGCTATTTGGCGGC 3561

Qy    521 LysAlaLysLysIleProSerGlyAlaSerPheArgGlyAspIleAsnValLeuLeuVal 540
Db   3562 ACGAATAAAACCTTTACCAAGGGTGGCCGTTATAGAGGTGACATAAATATTTTACTTTGT 3621

Qy    541 GlyAspProGlyThrSerLysSerGlnLeuLeuGlnTyrValHisLysIleAlaProArg 560
Db   3622 GGGGATCCTTCTACTTCCAAATCGCAAATTTTGCAATACGTTCACAAAATTACTCCTCGT 3681

Qy    561 GlyIleTyrThrSerGlyArgGlySerSerAlaValGlyLeuThrAlaTyrValThrLys 580
Db   3682 GGTGTGTATACTTCGGGTAAAGGTTCATCTGCCGTTGGTTTAACTGCTTATATTACAAGG 3741

Qy    581 AspProGluThrArgGluThrValLeuGluSerGlyAlaLeuValLeuSerAspArgGly 600
Db   3742 GATGTCGACACAAAACAACTTGTTTTGGAAAGTGGTGCATTAGTATTGTCTGATGGAGGT 3801

Qy    601 IleCysCysIleAspGluPheAspLysMetSerAspAsnAlaArgSerMetLeuHisGlu 620
Db   3802 GTTTGTTGTATTGACGAGTTTGATAAAATGAGTGATTCTACAAGGTCCGTCTTGCACGAA 3861

Qy    621 ValMetGluGlnGlnThrValSerValAlaLysGlyGlyIleIleAlaSerLeuAsnAla 640
Db   3862 GTCATGGAACAGCAGACTATTTCAATCGCAAAAGCGGGAATTATCACAACACTTAATGCC 3921

Qy    641 ArgThrSerValLeuAlaCysAlaAsnProSerGlySerArgTyrAsnAlaArgLeuSer 660
Db   3922 AGAAGTTCTATTTTGGCCAGTGCTAACCCAATTGGTTCACGCTACAACCCCAATTTGCCT 3981

Qy    661 ValIleAspAsnIleGlnLeuProProThrLeuLeuSerArgPheAspLeuIleTyrLeu 680
Db   3982 GTGACTGAAAATATTGATCTACCGCCCCCACTACTTTCGAGATTCGATCTGGTCTATCTT 4041

Qy    681 MetLeuAspLysProAspGluGlnAsnAspArgArgLeuAlaArgHisLeuValAlaLeu 700
Db   4042 GTCCTTGATAAGGTTGATGAGAAAAATGACAGAGAACTAGCCAAACACTTAACAAATCTT 4101

Qy    701 HisTyrGluAsn------TyrGluValSerLysGlnAspAlaLeuAspLeuGlnThrLeu 718
Db   4102 TACCTGGAAGATAAGCCCGAACATATTTCTCAAGACGACGTGCTACCAGTTGAATTTTTA 4161

Qy    719 ThrAlaTyrIleThrTyrAlaArgGlnHisValHisProThrLeuSerAspGluAlaAla 738
Db   4162 ACGATGTATATTAGTTATGCAAAGGAGCACATACACCCAATAATCACCGAGGCCGCTAAG 4221

Qy    739 GluAspLeuIleAsnGlyTyrValGluMetArgGlnLysGlyAsnPheProGlySerSer 758
Db   4222 ACTGAGCTTGTTCGTGCTTATGTAGGAATGAGAAAGATGGGTGACGATTCGAGATCCGAT 4281

Qy    759 LysLysValIleThrAlaThrProArgGlnLeuGluSerMetIleArgIleSerGluAla 778
Db   4282 GAGAAGAGAATCACAGCTACCACAAGACAACTTGAAAGTATGATTCGTTTGGCTGAGGCG 4341

Qy    779 LeuAlaArgMetArgPheSerGluValValGluLysValAspAlaAlaGluAlaValArg 798
Db   4342 CACGCCAAAATGAAATTGAAAAACGTCGTAGAGCTGGAGGATGTTCAAGAAGCCGTTAGA 4401

Qy    799 LeuLeuAspValAlaLeuGlnGlnSerAlaThrAspHisAlaThrGlyThrIleAspMet 818
Db   4402 TTAATTAGATCAGCCATAAAAGATTATGCAACAGACCCTAAAACCGGTAAAATCGACATG 4461

Qy    819 AspLeuIleThrThrGlyValSerAlaSerGluArgIleArgArgAlaAsnLeuLeuAla 838
Db   4462 AATTTAGTTCAAACAGGTAAATCAGTTATTCAGAGAAAACTACAGGAGGATTTGTCAAGG 4521

Qy    839 AlaLeuArgGluLeuIleAlaAspLysIleSerProGlySerSerSerGlyLeuLysThr 858
Db   4522 GAAATTATGAATGTTTTGAAGGATCAG------------GCATCAGACTCAATGTCATTC 4569

Qy    859 SerGlnLeuLeuGluAspIleArgSerGlnSerSerValAspValSerLeuGlnAspIle 878
Db   4570 AATGAGCTTATAAAACAAATCAATGAACACTCTCAAGATAGAGTTGAGTCTTCTGATATC 4629

Qy    879 LysAsnAlaLeuGlySerLeuGlnGlyGluGlyPheLeuThrValHisGlyAspIleVal 898
Db   4630 CAGGAAGCCTTGTCAAGATTGCAACAAGAGGACAAGGTCATTGTCCTTGGCGAGGGTGTA 4689

Qy    899 LysArg 900
Db   4690 AGGAGA 4695
```

RESULT 11
SC9531X
LOCUS   SC9531X   27166 bp   DNA   linear   PLN 18-APR-2005
DEFINITION S.cerevisiae chromosome XVI cosmid 9531.
ACCESSION   Z49919 U00094
VERSION   Z49919.1 GI:887584
KEYWORDS   CDC21 homologue; DSS4; F1Fo-ATP synthase g subunit; Grave's disease; guanine-releasing protein; Mitochondrial carrier protein; RNA polymerase I; RPA2; transfer RNA-Lys; zinc finger protein.

```
SOURCE      Saccharomyces cerevisiae (baker's yeast)
  ORGANISM  Saccharomyces cerevisiae
            Eukaryota; Fungi; Ascomycota; Saccharomycotina; Saccharomycetes;
            Saccharomycetales; Saccharomycetaceae; Saccharomyces.
REFERENCE   1  (bases 1 to 27166)
  AUTHORS   Bowman,S.
  JOURNAL   Unpublished
REFERENCE   2  (bases 1 to 27166)
  AUTHORS   Barrell,B.G., Rajandream,M.A. and Walsh,S.V.
  TITLE     Direct Submission
  JOURNAL   Submitted (23-JUN-1995) Saccharomyces cerevisiae chromosome XVI
            sequencing project, Sanger Centre, Hinxton Hall, Hinxton, Cambridge
            CB10 1RQ E-mail: barrell@sanger.ac.uk
COMMENT     Notes:
            All CDS over 100 codons have been analysed. CDS that are completely
            overlapped and those that are overlapped by more than 50%
            of their length by a larger CDS have been omitted from this
            analysis.
            Details of the omitted CDS are available on request. The more
            significant matches with motifs in the PROSITE database are
            also included but some of these may be fortuitous. The length in
            codons and the calculated codon adaptation index (CAI)
            is given for each CDS.
            Cosmid 9531 is overlapped at the 5' end by cosmid 9723, embl entry
            SC9723, accession no. Z48951, and at the 3' end by cosmid 9367,
            embl entry SC9367, accession no. Z49274.
FEATURES             Location/Qualifiers
     source          1. .27166
                     /organism="Saccharomyces cerevisiae"
                     /mol_type="genomic DNA"
                     /strain="AB972"
                     /db_xref="taxon:4932"
                     /chromosome="XVI"
                     /map="16R"
                     /clone="cosmid 9531"
     CDS             <1. .1727
                     /note="YP9531.01, unknown, len: 574, CAI: 0.14, overlaps
                     and extends cosmid 9723 orf YP9723.08 which has similarity
                     at N-terminus to transcription factors"
                     /codon_start=3
                     /product="unknown"
                     /protein_id="CAA90152.1"
                     /db_xref="GI:887585"
                     /db_xref="GOA:Q12753"
                     /db_xref="UniProt/Swiss-Prot:Q12753"
                     /translation="SISRSQPMFERVLSSTSLDSNMLSGHGALSDTSSILTSTFLDSE
                     PGVGKISKDYHHVPSLASISSLQSSQSLDQNFSIPQSPPLSSMSFNFLTGNINETNQN
                     HSNHQHSKSGNNWQDSSVSLPAKADSRLNMMDKNNSVGLDLLGHSKRISPISNSRVGE
                     VSVPLEEYIPSDIDGVGRVTDKSSLVYDWPFDESIERNFSTTATAATGESKFDINDNC
                     NRINSKSYSKTNSMNGNGMNNSNNNNINSNGNDKNNNNSSRQEHQGNGLFDMFTDSSS
                     ISTLSRANLLLQEKIGSQENSVKQENYSKNPQLRHQLTSRSRSFIHHPANEYLKNTFG
                     NSHSNDIGKGVEVLSLTPSFMDIPEKERETERSPSSNYITDRPFTRKPRSSSIDVNHR
                     YPPMAPTTVATSPGALNNAVASNLDDQLSLTSLNSQPSSIANMMMDPSNLAEQSSIHS
                     VPQSINSPRMPKTGSRQDKNIHTKKEERNPLNNIHDLSQLENVPDEMNQMFSPPLKSM
                     NRPDAIRENSSSSNFIIQGNSMISTPSGRNDLPDTSPMSSIQTASPPSQLLTDQGFAD
                     LDNFMSSL"
     misc_feature    1. .103
                     /note="overlap with cosmid 9723, positions 17558 to 17660;
                     EMBL:SC9723, Z48951"
     CDS             3177. .3983
                     /note="YP9531.02 unknown, len: , CAI: , possible zinc
                     finger protein"
                     /codon_start=1
                     /product="unknown"
                     /protein_id="CAA90153.1"
                     /db_xref="GI:887586"
                     /db_xref="GOA:Q12286"
                     /db_xref="SGD:S000006213"
                     /db_xref="UniProt/TrEMBL:Q12286"
                     /translation="MKPNNRTCDVITNKDESLPALLLPALNSYTCDDSLLKGQISSNG
                     RYQPFGFSDCSLLPKRLNIQAGQGSMPVSSIQCADHSYSNWQKESEKTKLPKLGCPTE
                     YTEYYKTVSSGETTDSAVVSSIATNRLKRKRQRDGPSCDSCRIKKIKCNATIIIFLQD
                     RNLISSISSNLHYTLSQDDINQFRMKFFRKLPDVMGTYEVIKHLDKIVLFKACTSCSR
                     RNQKNGKCLFSRGFTKSDMNVFPKINSKLKDKSIFEMTVDDYVAAGFQTL"
     gene            complement(4210. .7821)
                     /gene="RPA2"
     CDS             complement(4210. .7821)
                     /gene="RPA2"
                     /note="[9531.03c, RPA2 gene (RPA135, SRP3, RRN2)
                     SW:RPA2_YEAST P22138 DNA-directed RNA polymerase I 135 KD
                     polypeptide (EC 2.7.7.6) (A135)(RNA polymerase I subunit
                     2)"
                     /codon_start=1
                     /product="Rpa2p"
                     /protein_id="CAA90154.1"
                     /db_xref="GI:887587"
                     /db_xref="GOA:P22138"
                     /db_xref="InterPro:IPR007120"
                     /db_xref="InterPro:IPR007121"
                     /db_xref="InterPro:IPR007641"
                     /db_xref="InterPro:IPR007642"
                     /db_xref="InterPro:IPR007644"
                     /db_xref="InterPro:IPR007645"
                     /db_xref="InterPro:IPR007647"
                     /db_xref="InterPro:IPR009674"
                     /db_xref="SGD:S000006214"
                     /db_xref="UniProt/Swiss-Prot:P22138"
                     /translation="MSKVIKPPGQARTADFRTLERESRFINPPKDKSAFPLLQEAVQP
                     HIGSFNALTEGPDGGLLNLGVKDIGEKVIFDGKPLNSEDEISNSGYLGNKLSVSVEQV
                     SIAKPMSNDGVSSAVERKVYPSESRQRLTSYRGKLLLKLKWSVNNGEENLFEVRDCGG
                     LPVMLQSNRCHLNKMSPYELVQHKEESDEIGGYFIVNGIEKLIRMLIVQRRNHPMAII
                     RPSFANRGASYSHYGIQIRSVRPDQTSQTNVLHYLNDGQVTFRFSWRKNEYLVPVVMI
                     LKALCHTSDREIFDGIIGNDVKDSFLTDRLELLLRGFKKRYPHLQNRTQVLQYLGDKF
                     RVVFQASPDQSDLEVGQEVLDRIVLVHLGKDGSQDKFRMLLFMIRKLYSLVAGECSPD
                     NPDATQHQEVLLGGFLYGMILKEKIDEYLQNIIAQVRMDINRGMAINFKDKRYMSRVL
                     MRVNENIGSKMQYFLSTGNLVSQSGLDLQQVSGYTVVAEKINFYRFISHFRMVHRGSF
                     FAQLKTTTVRKLLPESWGFLCPVHTPDGSPCGLLNHFAHKCRISTQQSDVSRIPSILY
                     SLGVAPASHTFAAGPSLCCVQIDGKIIGWVSHEQGKIIADTLRYWKVEGKTPGLPIDL
                     EIGYVPPSTRGQYPGLYLFGGHSRMLRPVRYLPLDKEDIVGPFEQVYMNIAVTPQEIQ
                     NNVHTHVEFTPTNILSILANLTPFSDFNQSPRNMYQCQMGKQTMGTPGVALCHRSDNK
                     LYRLQTGQTPIVKANLYDDYGMDNFPNGFNAVVAVISYTGYDMDDAMIINKSADERGF
                     GYGTMYKTEKVDLALNRNRGDPITQHFGFGNDEWPKEWLEKLDEDGLPYIGTYVEEGD
                     PICAYFDDTLNKTKIKTYHSSEPAYIEEVNLIGDESNKFQELQTVSIKYRIRRTPQIG
                     DKFSSRHGQKGVCSRKWPTIDMPFSETGIQPDIIINPHAFPSRMTIGMFVESLAGKAG
                     ALHGIAQDSTPWIFNEDDTPADYFGEQLAKAGYNYHGNEPMYSGATGEELRADIYVGV
                     VYYQRLRHMVNDKFQVRSTGPVNSLTMQPVKGRKRHGGIRVGEMERDALIGHGTSFLL
                     QDRLLNSSDYTQASVCRECGSILTTQQSVPRIGSISTVCCRRCSMRFEDAKKLLTKSE
                     DGEKIFIDDSQIWEDGQGNKFVGGNETTTVAIPFVLKYLDSELSAMGIRLRYNVEPK"
     tRNA            8687. .8759
                     /product="tRNA-Lys"
                     /note="tRNA-Lys gene, anticodon ctt, len: 73"
     CDS             complement(9687. .10667)
                     /note="YP9531.04c, similar to SW:GDC_HUMAN P16260 Grave's
                     disease carrier protein (31.0% identity in 300 aa
                     overlap); contains 2 x PS00215 Mitochondrial energy
                     transfer proteins signature"
                     /codon_start=1
                     /product="unknown"
                     /protein_id="CAA90155.1"
                     /db_xref="GI:887588"
                     /db_xref="GOA:Q12251"
                     /db_xref="InterPro:IPR001993"
                     /db_xref="InterPro:IPR002067"
                     /db_xref="InterPro:IPR002113"
                     /db_xref="SGD:S000006215"
                     /db_xref="UniProt/TrEMBL:Q12251"
                     /translation="MAEVLTVLEQPNSIKDFLKQDSNIAFLAGGVAGAVSRTVVSPFE
                     RVKILLQVQSSTTSYNRGIFSSIRQVYHEEGTKGLFRGNGLNCIRIFPYSAVQFVVYE
                     ACKKKLFHVNGNNGQEQLTNTQRLFSGALCGGCSVVATYPLDLIKTRLSIQTANLSSL
                     NRSKAKSISKPPGIWQLLSETYRLEGGLRGLYRGVWPTSLGVVPYVALNFAVYEQLRE
                     FGVNSSDAQPSWKSNLYKLTIGAISGGVAQTITYPFDLLRRRFQVLAMGGNELGFRYT
                     SVWDALVTIGRAEGVSGYYKGLAANLFKVVPSTAVSWLVYEVVCDSVRNW"
     misc_feature    complement(10218. .10244)
                     /note="PS00215 Mitochondrial energy transfer proteins
                     signature"
     misc_feature    complement(10518. .10544)
```

```
                    /note="PS00215 Mitochondrial energy transfer proteins
                    signature"
     CDS            10934. .11188
                    /note="YP9531.05, unknown, len: 84, CAI: 0.12"
                    /codon_start=1
                    /product="unknown"
                    /protein_id="CAA90156.1"
                    /db_xref="GI:887589"
                    /db_xref="SGD:S000006216"
                    /db_xref="UniProt/TrEMBL:Q04203"
                    /translation="MELSQYLNYAFSLAYYIIIHLLCLSYIYEIIHKHKNVFVRPSKL
                    EDALPLYKTGKNTNKEGDLRLISFIPLLLKSVKHSIRVYT"
     CDS            complement(11257. .12210)
                    /note="YP9531.06c, unknown, len: 317, CAI: 0.13, putative
                    zinc finger protein, contains PS00028 Zinc finger, C2H2
                    type, domain"
                    /codon_start=1
                    /product="unknown"
                    /protein_id="CAA90157.1"
                    /db_xref="GI:887590"
                    /db_xref="GOA:Q12145"
                    /db_xref="InterPro:IPR007087"
                    /db_xref="SGD:S000006217"
                    /db_xref="UniProt/TrEMBL:Q12145"
                    /translation="MNFSLSKQSSEKQSSYTDKSRSPNIGMCTVNYKSNLPLQVSAVD
                    QLSTGKGTQFYGHNSYKNERECYNSTKINLPPISSLLPNFENNTPPNVDSRVQFPPQQ
                    VYQSMNVVPIVNEIYTPISMNATSDQYPIYYTESQQPIPHSQSPHLTSSAPLMMPVMV
                    PTVYKPLTPYDKEPITIASEPNFTAISMASHPNAALELCHDRPKSVPPGYGVLPTMQE
                    ASNGRTKSEPGAVLNGSATFSDWKTDTRISSTKLRKQCPVCGKICSRPSTLKTHYLIH
                    TGDTPFKCTWEGCTKSFNVKSNMLRHLKSHERKRNKVLNTT"
     misc_feature   complement(11293. .11361)
                    /note="PS00028 Zinc finger, C2H2 type, domain"
     misc_feature   complement(11383. .11445)
                    /note="PS00028 Zinc finger, C2H2 type, domain"
     CDS            complement(13814. .14143)
                    /note="YP9531.07c unknown, len: 109, CAI: 0.08"
                    /codon_start=1
                    /product="unknown"
                    /protein_id="CAA90158.1"
                    /db_xref="GI:887591"
                    /db_xref="SGD:S000006218"
                    /translation="MSCIFSADLGVEYSCAESRITNLVLCILCIREEKAAPVVKKDKF
                    LFFIFLSSKENLFCEISSRSSVSEVKSLRAKNNFICPHAVIGFQGAKAQLGIRFPQLN
                    GLLRLQH"
     CDS            complement(16908. .17651)
                    /note="YP9531.08c, unknown, len: 247, CAI: 0.11, putative
                    zinc finger protein, contains 2 x PS00028 Zinc finger,

Alignment Scores:
Pred. No.:              1e-58          Length:         27166
Score:                  1833.00        Matches:        393
Percent Similarity:     59.36%         Conservative:   178
Best Local Similarity:  40.85%         Mismatches:     281
Query Match:            39.77%         Indels:         110
DB:                     15             Gaps:           14

US-10-768-511-8 (1-901) x SC9531X (1-27166)

Qy          2 GluAsnAsnAspAlaLeuAspIleGlyAlaValSerSerProTyrProSerGlnSerGlu 21
Db      23405 GAGGATAATAACTCC------------------AGCTCCCCGTTGTGCCTAATCCTGAT 23446

Qy         22 GlyValSerThrProLeuProGlnValThrSerProSerPheAspAsnAlaAlaSerPro 41
Db      23447 TCTGTTCCA---------CCACAGCTTTCTTCCCCAGCTCTATTTTATAGCTCCTCTTCA 23497

Qy         42 ValAlaGly--------------------------------------------------- 44
Db      23498 TCACAAGGTGATATCTATGGTCGCAACAATAGCCAGAACTTAAGTCAGGGAGAGGGAAAC 23557

Qy         45 ---ArgArgAlaValArgGlnThrPro---------ThrSerAlaValArgArgArgGly 60
Db      23558 ATCAGAGCTGCTATAGGTTCTTCTCCACTAAATTTTCCATCTTCTTCCCAAAGACAAAAT 23617

Qy         61 ArgGluThrAspSerAlaArgArgArgArgSerArgSerArgSerLeuGlyAsnSerVal 80
Db      23618 TCCGATGTTTTCCAATCTCAAGGCAGACAGGGCAGAATTCGTTCTTCTGCGAGCGCTTCT 23677

Qy         81 TyrSerSerProTyrAspAlaGlyThrProGlyThrProGlyThrProValAlaThrPro 100
Db      23678 GGAAGGTCTAGATATCATTCTGATCTGAGAAGTGATAGAGCACTGCCTACTTCTTCC--- 23734

Qy        101 ValTyrAlaThrProValGlyThrProMetGlyThrProSerPheHisArgGlyThrPro 120
Db      23734 ------------------------------------------------------------ 23734

Qy        121 GlnTyrLysGlnArgSerGluLeuGlySerGlnGlyLysProLeuHisArgArgArgArg 140
Db      23735 ---------------TCTTCTTTAGGCCGTAATGGTCAAAACCGTGTACACATGCGAAGA 23779

Qy        141 SerGlnSerArgGluProGlyHisArgSerProSerArgGluProSerAlaAspGlyArg 160
Db      23780 AATGATATTCATACATCTGATTTATCCTCTCCAAGAAGAATTGTGGATTTTGATACTAGA 23839

Qy        161 ------------------------------------------ProSerGluSerAlaGlu 166
Db      23840 TCCGGCGTGAATACTTTGGATACTTCTTCTTCTTCTGCTCCTCCATCGGAAGCCAGTGAA 23899

Qy        167 ProAspAspThrLeuGlyGlyGluTyrAlaTyrValTrpGlyThrAsnValAsnIlePro 186
Db      23900 CCC---------------------TTGAGAATAATTTGGGGTACCAACGTCAGTATCCAG 23938

Qy        187 AspValLeuArgAlaIleArgArgPheLeuHisAsnTyrArg------------------ 200
Db      23939 GAATGTACAACTAATTTTCGTAATTTTTTGATGTCTTTCAAGTATAAATTCCGTAAAATA 23998

Qy        201 ------------------------SerSerAlaHisAspLeuAsnSerLysTyrIleGln 212
Db      23999 TTGGATGAAAGGGAGGAATTCATTAACAATACTACCGACGAAGAACTATACTATATCAAG 24058

Qy        213 IleIleGluGluThrValGluArgGluGluAspThrLeuAsnIleAspMetSerAspIle 232
Db      24059 CAGCTTAATGAAATGAGAGAACTTGGTACCTCTAATTTAAACTTGGATGCTAGAAACTTA 24118

Qy        233 ------TyrAspHisAspProAspLeuTyrAlaLysIleValArgTyrProLeuAspIle 250
Db      24119 CTTGCTTACAAGCAAACAGAAGACTTATACCATCAACTATTAAATTATCCTCAAGAAGTG 24178

Qy        251 IleProLeuLeuAspThrGluCysGlnGluValAlaThrSerLeu--------------- 265
Db      24179 ATTTCTATCATGGATCAAACTATCAAGGACTGTATGGTTTCCTTAATAGTAGACAACAAT 24238

Qy        266 ------------LeuProThrPheGlu---LysHisIleGluAlaArgProPheAsnLeu 280
Db      24239 TTGGATTACGATTTAGATGAAATAGAGACCAAGTTTTATAAAGTAAGGCCTTACAATGTA 24298

Qy        281 LysAlaSerValHisMetArgGluLeuAsnProSerAspIleAspLysLeuValSerVal 300
Db      24299 GGTTCTTGTAAAGGTATGCGCGAATTGAATCCAAATGATATTGATAAATTGATAAACTTA 24358

Qy        301 LysGlyMetValIleArgCysSerSerIleIleProGluIleLysGlyAlaPhePheLys 320
Db      24359 AAGGGTCTTGTCCTGAGATCAACTCCGGTTATCCCTGATATGAAGGTAGCGTTTTTCAAA 24418

Qy        321 CysLeuValCysGlyHisSerProProLeuValThrValValLysGlyArgValGluGlu 340
Db      24419 TGCAACGTCTGCGATCATACA---ATGGCAGTGGAAATTGATAGAGGAGTTATACAAGAG 24475

Qy        341 ProThrArgCysGluLysProGluCysAlaAlaArgAsnAlaMetSerLeuIleHisAsn 360
Db      24476 CCCGCTAGGTGTGAACGTATTGATTGTAATGAACCCAATTCCATGTCACTGATTCACAAC 24535

Qy        361 ArgCysThrPheAlaAsnLysGlnIleValArgLeuGlnGluThrProAspAlaIlePro 380
Db      24536 AGGTGTTCATTCGCAGATAAACAAGTCATTAAGTTACAGGAAACTCCAGACTTTGTGCCT 24595
```

```
Qy        381 GluGlyGluThrProHisThrValSerMetCysLeuTyrAsnThrMetValAspAlaVal 400
              :::|||:::||||||||::::::|||:::|||:::|||:::   :::||||||:::
Db      24596 GATGGACAAACGCCTCACTCTATCTCATTATGTGTTTACGATGAATTAGTGGATTCTTGC 24655

Qy        401 LysProGlyAspArgIleGluValThrGlyValPheLysAlaMetAlaValArgValGly 420
              :::      |||||||||||||||||||||||||   |||:::::::::   :::|||
Db      24656 AGGGCGGGCGATCGTATTGAGGTGACTGGCACGTTCAGGTCCATCCCCATTAGAGCTAAT 24715

Qy        421 ProAsnGlnArgThrLeuArgAlaLeuTyrLysThrTyrIleAspCysValHisValLys 440
                 ||||||   |||::::::||||||||||||||||:::|||   |||||||||||
Db      24716 TCCAGGCAACGCGTACTAAAGTCGTTGTATAAAACATACGTCGATGTGGTCCACGTTAAA 24775

Qy        441 LysSerAspArgGlyArgLeuGlnThrGluAspProMetGluMetAspLysGluAsnAsp 460
              |||            ||||||        :::             :::
Db      24776 AAAGTTTCAGATAAGAGGTTAGACGTCGATACTTCTACTATTGAACAAGAATTAATGCAG 24835

Qy        461 MetTyrAlaGlyTyrHisGluSerAspThrSerGluAlaAlaAsnGluAlaLysIleGln 480
                          ::::::|||   :::                  :::     :::
Db      24836 AACAAGGTAGATCATAACGAGGTCGAAGAAGTAAGACAAATTACTGATCAGGATTTAGCA 24895

Qy        481 LysLeuLysGluLeuSerLysLeuProGlyIleTyrAspArgLeuSerArgSerLeuAla 500
              |||::::::|||::::::             :::|||      |||:::|||||||:::|||
Db      24896 AAAATTCGCGAGGTTGCGGCAAGAGAAGATTTATACAGTTTATTAGCCCGCTCTATTGCC 24955

Qy        501 ProSerIleTrpGluLeuGluAspIleLysLysGlyLeuLeuCysGlnLeuPheGlyGly 520
              |||||||||:::|||||||||||||:::||||||:::|||   |||||||||||||||
Db      24956 CCAAGTATTTACGAGCTAGAAGATGTCAAGAAGGGTATATTACTTCAGCTATTTGGCGGC 25015

Qy        521 LysAlaLysLysIleProSerGlyAlaSerPheArgGlyAspIleAsnValLeuLeuVal 540
                       |||               :::||||||||||||||||:::||||||
Db      25016 ACGAATAAAACCTTTACCAAGGGTGGCCGTTATAGAGGTGACATAAATATTTTACTTTGT 25075

Qy        541 GlyAspProGlyThrSerLysSerGlnLeuLeuGlnTyrValHisLysIleAlaProArg 560
              ||||||||   ||||||||||||||:::||||||||||||||||||   ||||||
Db      25076 GGGGATCCTTCTACTTCCAAATCGCAAATTTTGCAATACGTTCACAAAATTACTCCTCGT 25135

Qy        561 GlyIleTyrThrSerGlyArgGlySerSerAlaValGlyLeuThrAlaTyrValThrLys 580
              |||:::||||||||||||:::|||||||||||||||||||||||||||||:::|||:::
Db      25136 GGTGTGTATACTTCGGGTAAAGGTTCATCTGCCGTTGGTTTAACTGCTTATATTACAAGG 25195

Qy        581 AspProGluThrArgGluThrValLeuGluSerGlyAlaLeuValLeuSerAspArgGly 600
              |||   :::|||::::::   |||||||||||||||||||||||||||||||   |||
Db      25196 GATGTCGACACAAAACAACTTGTTTTGGAAAGTGGTGCATTAGTATTGTCTGATGGAGGT 25255

Qy        601 IleCysCysIleAspGluPheAspLysMetSerAspAsnAlaArgSerMetLeuHisGlu 620
              :::|||||||||||||||||||||||||||||||||:::      ||||||:::||||||||
Db      25256 GTTTGTTGTATTGACGAGTTTGATAAAATGAGTGATTCTACAAGGTCCGTCTTGCACGAA 25315

Qy        621 ValMetGluGlnGlnThrValSerValAlaLysGlyGlyIleIleAlaSerLeuAsnAla 640
              ||||||||||||||||||:::|||:::||||||   ||||||||   :::||||||||
Db      25316 GTCATGGAACAGCAGACTATTTCAATCGCAAAAGCGGGAATTATCACAACACTTAATGCC 25375

Qy        641 ArgThrSerValLeuAlaCysAlaAsnProSerGlySerArgTyrAsnAlaArgLeuSer 660
              |||:::|||:::||||||   |||||||||   ||||||||||||||       |||
Db      25376 AGAAGTTCTATTTGGCCAGTGCTAACCCAATTGGTTCACGCTACAACCCCAATTTGCCT 25435

Qy        661 ValIleAspAsnIleGlnLeuProProThrLeuLeuSerArgPheAspLeuIleTyrLeu 680
              |||    :::||||||   |||||||||   ||||||||||||||||||||:::||||||
Db      25436 GTGACTGAAAATATTGATCTACCGCCCCCACTACTTTCGAGATTCGATCTGGTCTATCTT 25495

Qy        681 MetLeuAspLysProAspGluGlnAsnAspArgArgLeuAlaArgHisLeuValAlaLeu 700
              :::||||||||   ||||||:::|||||||||   ||||||:::|||||||     |||
Db      25496 GTCCTTGATAAGGTTGATGAGAAAAATGACAGAGAACTAGCCAAACACTTAACAAATCTT 25555

Qy        701 HisTyrGluAsn------TyrGluValSerLysGlnAspAlaLeuAspLeuGlnThrLeu 718
              :::   |||:::            :::|||:::   |||   |||   ::::::   |||
Db      25556 TACCTGGAAGATAAGCCCGAACATATTTCTCAAGACGACGTGCTACCAGTTGAATTTTTA 25615

Qy        719 ThrAlaTyrIleThrTyrAlaArgGlnHisValHisProThrLeuSerAspGluAlaAla 738
              |||   |||||:::||||||::::::|||:::||||||   :::::::::   |||
Db      25616 ACGATGTATATTAGTTATGCAAAGGAGCACATACACCCAATAATCACCGAGGCCGCTAAG 25675

Qy        739 GluAspLeuIleAsnGlyTyrValGluMetArgGlnLysGlyAsnPheProGlySerSer 758
              :::|||:::      ||||||   ||||||:::   |||:::          |||
Db      25676 ACTGAGCTTGTTCGTGCTTATGTAGGAATGAGAAAGATGGGTGACGATTCGAGATCCGAT 25735

Qy        759 LysLysValIleThrAlaThrProArgGlnLeuGluSerMetIleArgIleSerGluAla 778
              :::|||   ||||||||||||   |||||||||||||||||||||||||::::::||||||
Db      25736 GAGAAGAGAATCACAGCTACCACAAGACAACTTGAAAGTATGATTCGTTTGGCTGAGGCG 25795

Qy        779 LeuAlaArgMetArgPheSerGluValValGluLysValAspAlaAlaGluAlaValArg 798
                 |||:::|||:::         ||||||||      |||      |||||||||||
Db      25796 CACGCCAAAATGAAATTGAAAAACGTCGTAGAGCTGGAGGATGTTCAAGAAGCCGTTAGA 25855

Qy        799 LeuLeuAspValAlaLeuGlnGlnSerAlaThrAspHisAlaThrGlyThrIleAspMet 818
              |||:::      |||::::::      ||||||||      ||||||   ||||||||
Db      25856 TTAATTAGATCAGCCATAAAAGATTATGCAACAGACCCTAAAACCGGTAAAATCGACATG 25915

Qy        819 AspLeuIleThrThrGlyValSerAlaSerGluArgIleArgArgAlaAsnLeuLeuAla 838
              :::|||:::   ||||||   |||      :::|||      :::   :::|||
Db      25916 AATTTAGTTCAAACAGGTAAATCAGTTATTCAGAGAAAACTACAGGAGGATTTGTCAAGG 25975

Qy        839 AlaLeuArgGluLeuIleAlaAspLysIleSerProGlySerSerGlyLeuLysThr 858
                 :::   ::::::   |||:::                :::|||      :::
Db      25976 GAAATTATGAATGTTTTGAAGGATCAG------------GCATCAGACTCAATGTCATTC 26023

Qy        859 SerGlnLeuLeuGluAspIleArgSerGlnSerSerValAspValSerLeuGlnAspIle 878
              ::::::|||::::::   |||         |||         |||         ||||||
Db      26024 AATGAGCTTATAAAACAAATCAATGAACACTCTCAAGATAGAGTTGAGTCTTCTGATATC 26083

Qy        879 LysAsnAlaLeuGlySerLeuGlnGlyGluGlyPheLeuThrValHisGlyAspIleVal 898
              :::   ||||||      ||||||   |||      :::   |||   |||:::   |||
Db      26084 CAGGAAGCCTTGTCAAGATTGCAACAAGAGGACAAGGTCATTGTCCTTGGCGAGGGTGTA 26143

Qy        899 LysArg 900
              :::|||
Db      26144 AGGAGA 26149
```

RESULT 12
SCCHRXVI

```
LOCUS       SCCHRXVI               165536 bp    DNA     linear   PLN 18-APR-2005
DEFINITION  S.cerevisiae chromosome XVI 165536 bp sequence, cen rightwards.
ACCESSION   Z71255
VERSION     Z71255.1  GI:1279666
KEYWORDS    ARO7; ATPase; beta-transducin; CCL1; CDC21 homologue; CDE1; CDE2;
            CDE3; centromere; CHL1; chorismate mutase; citrate synthase;
            coatomer zeta subunit; cyclin; delta element; DnaJ; DSS4; electron
            transport flavoprotein; eukaryotic iniation factor 5; F1Fo-ATP
            synthase g subunit; gamma adaptin; GLN1; glutamine synthetase;
            Grave's disease; guanine-releasing protein; HAL1; halotolerance;
            histidine tRNA synthetase; HTS1; isocitrate lyase; LPT1; LTR; MAK3;
            MFS1; Mitochondrial carrier protein; N-acetyltransferase;
            NADPH-cytochrome P450 reductase; negative regulator; NHP6A;
            nonhistone chromosomal protein 6A; OSD1; phenylalanyl-tRNA
            synthetase; phosphotyrosine protein phosphatase; polyposis;
            protamine P1; protein kinase; protein transport; regulatory
            protein; ribosomal protein L37a; RNA polymerase I; ROX1; RPA2;
            SEC8; SMK1; spermidine synthase; TIF5; TKL1; transfer RNA-Gly;
            transfer RNA-Lys; transfer RNA-Phe; transfer RNA-Ser;
            transketolase; tropomyosin; ubiquitin-activating enzyme; vacuolar
            H(+)-ATPase 54 kda subunit; VMA13; YMC1; YME1; YTA11; zinc finger
            protein.
SOURCE      Saccharomyces cerevisiae (baker's yeast)
  ORGANISM  Saccharomyces cerevisiae
            Eukaryota; Fungi; Ascomycota; Saccharomycotina; Saccharomycetes;
            Saccharomycetales; Saccharomycetaceae; Saccharomyces.
REFERENCE   1  (bases 1 to 165536)
  AUTHORS   Badcock,K., Bowman,S., Churcher,C.M., Pearson,D., Rajandream,M.A.,
            Walsh,S.V. and Barrell,B.G.
  TITLE     Direct Submission
  JOURNAL   Submitted (19-APR-1996) Saccharomyces cerevisiae chromosome IV
            sequencing project, Sanger Centre, Hinxton Hall, Hinxton, Cambridge
            CB10 1RQ. E-mail: barrell@sanger.ac.uk
COMMENT     SC8132X   Z48483       1 to 25540
            SC9723    Z48951     100 to 17660
            SC9531X   Z49919     104 to 27166
```

```
            SC9367   Z49274      10 to 42190
            SC3085   Z68111     790 to 10339   (PCR product)
            SC9499X  Z49219     134 to 43776
            The most significant matches using FASTA -o to a non-redundant
            database compiled from Swissprot 32 and PIR 46 and Wormpep9 are
            included.
            All CDS over 100 codons have been analysed. CDS that are largely or
            completely overlapped by a larger CDS have been omitted from this
            analysis.
            The more significant matches with motifs in the PROSITE database
            are also included but some of these may be fortuitous. The length
            in codons/amino-acids is given for each CDS as is the calculated
            codon adaptation index (CAI).
            tRNA genes and LTRs are also included.
            Notes:
            This sequence has been compiled from the following submissions of
            individual contigs.
FEATURES             Location/Qualifiers
     source          1. .165536
                     /organism="Saccharomyces cerevisiae"
                     /mol_type="genomic DNA"
                     /strain="AB972"
                     /db_xref="taxon:4932"
                     /chromosome="XVI"
     gene            <1. .3032
                     /gene="YPL012W"
     CDS             <1. .3032
                     /gene="YPL012W"
                     /note="YPL012W, unknown, incomplete, len: 1009, CAI: 0.19"
                     /codon_start=3
                     /product="unknown"
                     /protein_id="CAA95029.1"
                     /db_xref="GI:1314064"
                     /db_xref="GOA:Q7LH12"
                     /db_xref="InterPro:IPR000357"
                     /db_xref="UniProt/TrEMBL:Q7LH12"
                     /translation="SNLSNKQLKAQKTKEDINASVMRSLRLITSVVSTGQWPSSQIEP
                     LCDVLLGVTKSSEQYLVSASFECFESMFKTMAETTISSGLAENKYLRVLDTIFALKPS
                     NVDTLLTKSWIAVVIKGMSTYATHQPLKALRKIPGVFHIMCTYLASETPEVYQAASQC
                     LISILSESVKDDLLLYTPSVDEKVFKNVDEIISQIAKTFIDFLSIRYSHCSREILKIL
                     VAAFNKFRYRSNPHFLKSLKIVDTWRVNEEQFMDLRNEIELVIGASISAMGPEMILAE
                     APLNLDNPSSERPGRAWLLPLIRDYTKNANLATFQNELAPYIKSFQSKFDKVPEESIQ
                     LRVFQTIVDQIWSTLPRFCELPMDLRESFTDEFASELSSLLYSEVELRTTICHALKVL
                     AESNVSYAEESSSHNVLLLQRFPISEAQKNIEYLSTKSTNLLAVLFNVYTQTTPNARS
                     YILETIDQYLKITSKEDLEKTFNNVCGLLKNSMNEESSGNVNKEKKPQLTATLLDLI
                     ICMITYLPVSSYSALFSMFSLTVNSADALIQKRAYRIITKLSELKSGSTAVAQFISDI
                     ENVMVDSASSVQTSAKAARLTAIKTIVELLPLDHLDFIVRTVAEVILSTKDVNEKSRE
                     TAFDTLICMGRKMNEPNGIIKLFQIPGYDPTTPDQSSSISEFFKIISAGLIGESQHMV
                     SSSITGYACLVFEFKNELDSGILMDIYDTIELYLTSNSREIVKSAIGFTKVCVLGLPE
                     ELMRPKVPELLLKLLRWSHEHTGHFKAKVKHIIERLIRRFGYDYIEANFPEEDRRLLT
                     NIRKMRNRNKRKDEEVTTGVSDVAATKGSRFMSAFDEAVYGSDEENDNGSDQEENVAG
                     GKMKNGAKQFIVESGDNPLDLLDSQTLAHISSTRPKKFNKNQNRARFNDDAFNFDSEG
                     KLVVKGQPKPSTNVDDPLSAVTSGINAYLEAVKSGPVRGQRNKLKFRKNGKDSDEFGD
                     DDDGEKDSRLMRGRVNQGNKIGKHNKKGPKFKSRKKL"
     gene            complement(3265. .4326)
                     /gene="YPL011C"
     CDS             complement(3265. .4326)
                     /gene="YPL011C"
                     /note="YPL011C, unknown, len: 353, CAI: 0.12"
                     /codon_start=1
                     /product="unknown"
                     /protein_id="CAA95030.1"
                     /db_xref="GI:1314065"
                     /db_xref="GOA:Q12297"
                     /db_xref="InterPro:IPR006565"
                     /db_xref="InterPro:IPR009072"
                     /db_xref="SGD:S000005932"
                     /db_xref="UniProt/Swiss-Prot:Q12297"
                     /translation="MTTNNDFYFALLRISILQLLKAQGFDRARPSLVDVMTDLYAKFL
                     SLLASEVSSIAQARCDQDDTIALQDITLALENLGIVKPTNVLDVYDENSELSSSRGME
                     KFKDWCIYSTQLTDARITALPTVELLQSEEKESDPLSAIPDYLNQLLQNKGAKQKLET
                     KNRKTELIEDLINNNGLDDWIKLVIARQRINMIERASKKESQNVPALPHIAGYKSSIL
                     SRHHHTTITNEDRMPSAMTPRDEDALTEIQENPFVTSKLPIMRKENRLENITLSFEDE
                     ELESLGEVEGPNQKSQENNNEESFKENNKSLTESPHGDDRDISMFQFDSNVDTKWAEQ
                     EDMDSTFQRRTSLDYGGYF"
     gene            4641. .5210
                     /gene="YPL010W"
     CDS             4641. .5210
                     /gene="YPL010W"
                     /note="YPL010W, unknown, len: 189, CAI: 0.19, 39.7%
                     identity in 156 aa overlap to coatomer zeta subunit
                     COPZ_BOVIN P35604"
                     /codon_start=1
                     /product="unknown"
                     /protein_id="CAA95031.1"
                     /db_xref="GI:1314066"
                     /db_xref="GOA:P53600"
                     /db_xref="InterPro:IPR000804"
                     /db_xref="InterPro:IPR011012"
                     /db_xref="SGD:S000005931"
                     /db_xref="UniProt/Swiss-Prot:P53600"
                     /translation="MSSLSLYTVQAVLILDQQGERIYAKYYQPPHRSDEGHQLLFNSV
                     KKQKEFEKQLYRKTHKQDSEILIFEDHLVLYKEYIDITIYLVASLEENEIVLQQGFSA
                     IRGALDLILNSGMDKKNIQENYDMVLLAIDETIDNGVILETDSNTIASRVSKPPTNEP
                     QMALDLDKGFLGAWGFAKSKFQERLQQGL"
     gene            complement(5443. .8559)
                     /gene="YPL009C"
     CDS             complement(5443. .8559)
                     /gene="YPL009C"
                     /note="YPL009C, unknown, len: 1038, CAI: 0.19, contains
                     glutamic acid rich region"
                     /codon_start=1
                     /product="unknown"
                     /protein_id="CAA95032.1"
                     /db_xref="GI:1314067"
                     /db_xref="GOA:Q12532"
                     /db_xref="InterPro:IPR008532"
                     /db_xref="SGD:S000005930"
                     /db_xref="UniProt/TrEMBL:Q12532"
                     /translation="MKQRISALDLLLLARELKQDLEGYRLSNIYNIADSSKQFLLKFN
                     KPDSKLNVVVDCGLRIYLTEFSRPIPPTPSGFVVKLRKHLKAKRLTALKQVDQDRILV
                     LQFADGHFYLVLEFFSAGNVILLDENRRIMALQRVVLEHENKVGQIYEMFDESLFTTN
                     NESADESIEKNRKAEYTSELVNEWIKAVQAKYESDITVIKQLNIQGKEGAKKKKVKVP
                     SIHKLLLSKVPHLSSDLLSKNLKVFNIDPSESCLNLLEETDSLAELLNSTQLEYNQLL
                     TTTDRKGYILAKRNENYISEKDTADLEFIYDTFHPFKPYINGGDTDSSCIIEVEGPYN
                     RTLDKFFSTIESSKYALRIQNQESQAQKKIDDARAENDRKIQALLDVQELNERKGHLI
                     IENAPLIEEVKLAVQGLIDQQMDWNTIEKLIKSEQKKGNRIAQLLNLPLNLKQNKISV
                     KLDLSSKELNTSSDEDNESEGNTTDSSSDSDSEDMESSKERSTKSMKRKSNEKINVTI
                     DLGLSAYANATEYFNIKKTSAQKQKKVEKNVGKAMKNIEVKIDQQLKKKLKDSHSVLK
                     KIRTPYFFEKYSWFISSEGFLVMMGKSPAETDQIYSKYIEDDDIYMSNSFNSHVWIKN
                     PEKTEVPPNTLMQAGILCMSSSEAWSKKISSSPWWCFAKNVSKFDGSDNSILPEGAFR
                     LKNENDQNHLPPAQLVMGFGFLWKVKTSGNEDNGDDDEEEEEEEEEEEEEEEEEEEEE
                     EEEKEEEEKEEEQQQDEDDSNEVNGLEKGGDSNDSTKNNSFEHDNLEKDIEKHCTISS
                     DTDSDSGNAKAKNDNSSTQRILDEPGVPISLIENINSNVRGKRGKLKKIQKKYADQDE
                     TERLLRLEALGTLKGIEKQQQRKKEEIMKREVREDRKNKREKQRRLQALKFTKKEKAR
                     VNYDKHKSELKPSLDKGDVVDDIIPVFAPWPALLKYKYKVKIQPGSAKKTKTLTEILH
                     YFKSRPLDGSSTDNEMDWPQEHEMIKGLKEQDLVLLLCVDKLKVTIAGQKSTKNGGNS
                     SKKGKKKR"
     gene            9008. .11593
                     /gene="CHL1"
     CDS             9008. .11593
                     /gene="CHL1"
                     /note="YPL008W, CHL1 gene, len: 861 CAI: 0.11
                     SW:CHL1_YEAST P22516, CHL1 protein, contains PS00690
                     DEAH-box subfamily ATP-dependent helicases signature"
                     /codon_start=1
                     /product="Chl1p"
                     /protein_id="CAA95033.1"
                     /db_xref="GI:1314068"
                     /db_xref="GOA:P22516"
                     /db_xref="InterPro:IPR001945"
                     /db_xref="InterPro:IPR002464"
                     /db_xref="InterPro:IPR006554"
                     /db_xref="InterPro:IPR006555"
                     /db_xref="InterPro:IPR010614"
                     /db_xref="SGD:S000005929"
```

```
                /db_xref="UniProt/Swiss-Prot:P22516"
                /translation="MDKKEYSETFYHPYKPYDIQVQLMETVYRVLSEGKKIAILESPT
                GTGKTLSLICATMTWLRMNKADIFTRMETNIKTNEDDSENLSDDEPDWVIDTYRKSVL
                QEKVDLLNDYEKHLNEINTTSCKQLKTMCDLDKEHGRYKSVDPLRKKRKGARHLDVSL
                EEQDFIPRPYESDSENNDTSKSTRGGRISDKDYKLSELNSQIITLLDKIDGKVSRDPN
                NGDRFDVTNQNPVKIYYASRTYSQLGQFTSQLRLPSFPSSFRDKVPDEKVKYLPLASK
                KQLCINPKVMKWKTLEAINDACADLRHSKEGCIFYQNTNEWRHCPDTLALRDMIFSEI
                QDIEDLVPLGKSLGICPYYASREALPIAEVVTLPYQYLLSESTRSSLQINLENSIVII
                DEAHNLIETINSIYSSQISLEDLKNCHKGIVTYFNKFKSRLNPGNRVNLLKLNSLLMT
                LIQFIVKNFKKIGQEIDPNDMFTGSNIDTLNIHKLLRYIKVSKIAYKIDTYNQALKEE
                ESSKNENPIKETHKKSVSSQPLLFKVSQFLYCLTNLTSEGQFFFEKNYSIKYMLLEPS
                KPFESILNQAKCVVLAGGTMEPMSEFLSNLLPEVPSEDITTLSCNHVIPKENLQTYIT
                NQPELEFTFEKRMSPSLVNNHLFQFFVDLSKAVPKKGGIVAFFPSYQYLAHVIQCWKQ
                NDRFATLNNVRKIFYEAKDGDDILSGYSDSVAEGRGSLLLAIVGGKLSEGINFQDDLC
                RAVVMVGLPFPNIFSGELIVKRKHLAAKIMKSGGTEEEASRATKEFMENICMKAVNQS
                VGRAIRHANDYANIYLLDVRYNRPNFRKKLSRWVQDSINSEHTTHQVISSTRKFFSMR
                SLNSR"
     misc_feature   10169. .10198

Alignment Scores:
Pred. No.:                6.01e-58      Length:         165536
Score:                    1833.00       Matches:        393
Percent Similarity:       59.36%        Conservative:   178
Best Local Similarity:    40.85%        Mismatches:     281
Query Match:              39.77%        Indels:         110
DB:                       15            Gaps:           14

US-10-768-511-8 (1-901) x SCCHRXVI (1-165536)

Qy          2 GluAsnAsnAspAlaLeuAspIleGlyAlaValSerSerProTyrProSerGlnSerGlu 21
Db      66403 GAGGATAATAACTCC------------------AGCTCCCCCGTTGTGCCTAATCCTGAT 66444

Qy         22 GlyValSerThrProLeuProGlnValThrSerProSerPheAspAsnAlaAlaSerPro 41
Db      66445 TCTGTTCCA---------CCACAGCTTTCTTCCCCAGCTCTATTTTATAGCTCCTCTTCA 66495

Qy         42 ValAlaGly--------------------------------------------------- 44
Db      66496 TCACAAGGTGATATCTATGGTCGCAACAATAGCCAGAACTTAAGTCAGGGAGAGGGAAAC 66555

Qy         45 ---ArgArgAlaValArgGlnThrPro---------ThrSerAlaValArgArgArgGly 60
Db      66556 ATCAGAGCTGCTATAGGTTCTTCTCCACTAAATTTTCCATCTTCTTCCCAAAGACAAAAT 66615

Qy         61 ArgGluThrAspSerAlaArgArgArgArgSerArgSerArgSerLeuGlyAsnSerVal 80
Db      66616 TCCGATGTTTTCCAATCTCAAGGCAGACAGGGCAGAATTCGTTCTTCTGCGAGCGCTTCT 66675

Qy         81 TyrSerSerProTyrAspAlaGlyThrProGlyThrProGlyThrProValAlaThrPro 100
Db      66676 GGAAGGTCTAGATATCATTCTGATCTGAGAAGTGATAGAGCACTGCCTACTTCTTCC--- 66732

Qy        101 ValTyrAlaThrProValGlyThrProMetGlyThrProSerPheHisArgGlyThrPro 120
Db      66732 ------------------------------------------------------------ 66732

Qy        121 GlnTyrLysGlnArgSerGluLeuGlySerGlnGlyLysProLeuHisArgArgArgArg 140
Db      66733 ---------------TCTTCTTTAGGCCGTAATGGTCAAAACCGTGTACACATGCGAAGA 66777

Qy        141 SerGlnSerArgGluProGlyHisArgSerProSerArgGluProSerAlaAspGlyArg 160
Db      66778 AATGATATTCATACATCTGATTTATCCTCTCCAAGAAGAATTGTGGATTTTGATACTAGA 66837

Qy        161 ------------------------------------------ProSerGluSerAlaGlu 166
Db      66838 TCCGGCGTGAATACTTTGGATACTTCTTCTTCTTCTGCTCCTCCATCGGAAGCCAGTGAA 66897

Qy        167 ProAspAspThrLeuGlyGlyGluTyrAlaTyrValTrpGlyThrAsnValAsnIlePro 186
Db      66898 CCC---------------------TTGAGAATAATTTGGGGTACCAACGTCAGTATCCAG 66936

Qy        187 AspValLeuArgAlaIleArgArgPheLeuHisAsnTyrArg------------------ 200
Db      66937 GAATGTACAACTAATTTTCGTAATTTTTTGATGTCTTTCAAGTATAAATTCCGTAAAATA 66996

Qy        201 ------------------------SerSerAlaHisAspLeuAsnSerLysTyrIleGln 212
Db      66997 TTGGATGAAAGGGAGGAATTCATTAACAATACTACCGACGAAGAACTATACTATATCAAG 67056

Qy        213 IleIleGluGluThrValGluArgGluGluAspThrLeuAsnIleAspMetSerAspIle 232
Db      67057 CAGCTTAATGAAATGAGAGAACTTGGTACCTCTAATTTAAACTTGGATGCTAGAAACTTA 67116

Qy        233 ------TyrAspHisAspProAspLeuTyrAlaLysIleValArgTyrProLeuAspIle 250
Db      67117 CTTGCTTACAAGCAAACAGAAGACTTATACCATCAACTATTAAATTATCCTCAAGAAGTG 67176

Qy        251 IleProLeuLeuAspThrGluCysGlnGluValAlaThrSerLeu--------------- 265
Db      67177 ATTTCTATCATGGATCAAACTATCAAGGACTGTATGGTTTCCTTAATAGTAGACAACAAT 67236

Qy        266 ------------LeuProThrPheGlu---LysHisIleGluAlaArgProPheAsnLeu 280
Db      67237 TTGGATTACGATTTAGATGAAATAGAGACCAAGTTTTATAAAGTAAGGCCTTACAATGTA 67296

Qy        281 LysAlaSerValHisMetArgGluLeuAsnProSerAspIleAspLysLeuValSerVal 300
Db      67297 GGTTCTTGTAAAGGTATGCGCGAATTGAATCCAAATGATATTGATAAATTGATAAACTTA 67356

Qy        301 LysGlyMetValIleArgCysSerSerIleIleProGluIleLysGlyAlaPhePheLys 320
Db      67357 AAGGGTCTTGTCCTGAGATCAACTCCGGTTATCCCTGATATGAAGGTAGCGTTTTTCAAA 67416

Qy        321 CysLeuValCysGlyHisSerProProLeuValThrValValLysGlyArgValGluGlu 340
Db      67417 TGCAACGTCTGCGATCATACA---ATGGCAGTGGAAATTGATAGAGGAGTTATACAAGAG 67473

Qy        341 ProThrArgCysGluLysProGluCysAlaAlaArgAsnAlaMetSerLeuIleHisAsn 360
Db      67474 CCCGCTAGGTGTGAACGTATTGATTGTAATGAACCCAATTCCATGTCACTGATTCACAAC 67533

Qy        361 ArgCysThrPheAlaAsnLysGlnIleValArgLeuGlnGluThrProAspAlaIlePro 380
Db      67534 AGGTGTTCATTCGCAGATAAACAAGTCATTAAGTTACAGGAAACTCCAGACTTTGTGCCT 67593

Qy        381 GluGlyGluThrProHisThrValSerMetCysLeuTyrAsnThrMetValAspAlaVal 400
Db      67594 GATGGACAAACGCCTCACTCTATCTCATTATGTGTTTACGATGAATTAGTGGATTCTTGC 67653

Qy        401 LysProGlyAspArgIleGluValThrGlyValPheLysAlaMetAlaValArgValGly 420
Db      67654 AGGGCGGGCGATCGTATTGAGGTGACTGGCACGTTCAGGTCCATCCCCATTAGAGCTAAT 67713

Qy        421 ProAsnGlnArgThrLeuArgAlaLeuTyrLysThrTyrIleAspCysValHisValLys 440
Db      67714 TCCAGGCAACGCGTACTAAAGTCGTTGTATAAAACATACGTCGATGTGGTCCACGTTAAA 67773

Qy        441 LysSerAspArgGlyArgLeuGlnThrGluAspProMetGluMetAspLysGluAsnAsp 460
Db      67774 AAAGTTTCAGATAAGAGGTTAGACGTCGATACTTCTACTATTGAACAAGAATTAATGCAG 67833

Qy        461 MetTyrAlaGlyTyrHisGluSerAspThrSerGluAlaAlaAsnGluAlaLysIleGln 480
Db      67834 AACAAGGTAGATCATAACGAGGTCGAAGAAGTAAGACAAATTACTGATCAGGATTTAGCA 67893

Qy        481 LysLeuLysGluLeuSerLysLeuProGlyIleTyrAspArgLeuSerArgSerLeuAla 500
Db      67894 AAAATTCGCGAGGTTGCGGCAAGAGAAGATTTATACAGTTTATTAGCCCGCTCTATTGCC 67953

Qy        501 ProSerIleTrpGluLeuGluAspIleLysLysGlyLeuLeuCysGlnLeuPheGlyGly 520
Db      67954 CCAAGTATTTACGAGCTAGAAGATGTCAAGAAGGGTATATTACTTCAGCTATTTGGCGGC 68013

Qy        521 LysAlaLysLysIleProSerGlyAlaSerPheArgGlyAspIleAsnValLeuLeuVal 540
```

```
                           |||          |||       :::||||||||||||||:::||||||
Db     68014 ACGAATAAAACCTTTACCAAGGGTGGCCGTTATAGAGGTGACATAAATATTTTACTTTGT 68073

Qy       541 GlyAspProGlyThrSerLysSerGlnLeuLeuGlnTyrValHisLysIleAlaProArg 560
             |||||||||   ||||||||||||||:::||||||||||||||||||||   ||||||
Db     68074 GGGGATCCTTCTACTTCCAAATCGCAAATTTTGCAATACGTTCACAAAATTACTCCTCGT 68133

Qy       561 GlyIleTyrThrSerGlyArgGlySerSerAlaValGlyLeuThrAlaTyrValThrLys 580
             |||:::||||||||||||:::||||||||||||||||||||||||||:::|||:::
Db     68134 GGTGTGTATACTTCGGGTAAAGGTTCATCTGCCGTTGGTTTAACTGCTTATATTACAAGG 68193

Qy       581 AspProGluThrArgGluThrValLeuGluSerGlyAlaLeuValLeuSerAspArgGly 600
             |||   :::|||::::::   ||||||||||||||||||||||||||||||   |||
Db     68194 GATGTCGACACAAAACAACTTGTTTTGGAAAGTGGTGCATTAGTATTGTCTGATGGAGGT 68253

Qy       601 IleCysCysIleAspGluPheAspLysMetSerAspAsnAlaArgSerMetLeuHisGlu 620
             :::||||||||||||||||||||||||||||||:::   ||||||:::|||||||||
Db     68254 GTTTGTTGTATTGACGAGTTTGATAAAATGAGTGATTCTACAAGGTCCGTCTTGCACGAA 68313

Qy       621 ValMetGluGlnGlnThrValSerValAlaLysGlyGlyIleIleAlaSerLeuAsnAla 640
             ||||||||||||||||||:::|||:::||||||   |||||||||   :::|||||||||
Db     68314 GTCATGGAACAGCAGACTATTTCAATCGCAAAAGCGGGAATTATCACAACACTTAATGCC 68373

Qy       641 ArgThrSerValLeuAlaCysAlaAsnProSerGlySerArgTyrAsnAlaArgLeuSer 660
             |||:::|||:::||||||   |||||||||   ||||||||||||||||   |||
Db     68374 AGAAGTTCTATTTTGGCCAGTGCTAACCCAATTGGTTCACGCTACAACCCCAATTTGCCT 68433

Qy       661 ValIleAspAsnIleGlnLeuProProThrLeuLeuSerArgPheAspLeuIleTyrLeu 680
             |||   :::||||||   |||||||||   |||||||||||||||||||:::||||||
Db     68434 GTGACTGAAAATATTGATCTACCGCCCCCACTACTTTCGAGATTCGATCTGGTCTATCTT 68493

Qy       681 MetLeuAspLysProAspGluGlnAsnAspArgArgLeuAlaArgHisLeuValAlaLeu 700
             :::|||||||||   ||||||:::|||||||||   ||||||:::||||||      |||
Db     68494 GTCCTTGATAAGGTTGATGAGAAAAATGACAGAGAACTAGCCAAACACTTAACAAATCTT 68553

Qy       701 HisTyrGluAsn------TyrGluValSerLysGlnAspAlaLeuAspLeuGlnThrLeu 718
             :::   |||:::         :::|||:::   |||   |||   ::::::   |||
Db     68554 TACCTGGAAGATAAGCCCGAACATATTTCTCAAGACGACGTGCTACCAGTTGAATTTTTA 68613

Qy       719 ThrAlaTyrIleThrTyrAlaArgGlnHisValHisProThrLeuSerAspGluAlaAla 738
             |||   ||||||:::||||||::::::|||:::||||||   :::::::::   |||
Db     68614 ACGATGTATATTAGTTATGCAAAGGAGCACATACACCCAATAATCACCGAGGCCGCTAAG 68673

Qy       739 GluAspLeuIleAsnGlyTyrValGluMetArgGlnLysGlyAsnPheProGlySerSer 758
             :::|||:::      ||||||   ||||||:::   |||:::         |||
Db     68674 ACTGAGCTTGTTCGTGCTTATGTAGGAATGAGAAAGATGGGTGACGATTCGAGATCCGAT 68733

Qy       759 LysLysValIleThrAlaThrProArgGlnLeuGluSerMetIleArgIleSerGluAla 778
             :::|||   ||||||||||||   ||||||||||||||||||||||||::::::||||||
Db     68734 GAGAAGAGAATCACAGCTACCACAAGACAACTTGAAAGTATGATTCGTTTGGCTGAGGCG 68793

Qy       779 LeuAlaArgMetArgPheSerGluValValGluLysValAspAlaAlaGluAlaValArg 798
                |||:::|||:::         |||||||||      |||      ||||||||||||
Db     68794 CACGCCAAAATGAAATTGAAAAACGTCGTAGAGCTGGAGGATGTTCAAGAAGCCGTTAGA 68853

Qy       799 LeuLeuAspValAlaLeuGlnGlnSerAlaThrAspHisAlaThrGlyThrIleAspMet 818
             |||:::         |||:::::::         |||||||||      ||||||   |||||||||
Db     68854 TTAATTAGATCAGCCATAAAAGATTATGCAACAGACCCTAAAACCGGTAAAATCGACATG 68913

Qy       819 AspLeuIleThrThrGlyValSerAlaSerGluArgIleArgArgAlaAsnLeuLeuAla 838
             :::|||:::   ||||||   |||      :::|||      :::   :::|||
Db     68914 AATTTAGTTCAAACAGGTAAATCAGTTATTCAGAGAAAACTACAGGAGGATTTGTCAAGG 68973

Qy       839 AlaLeuArgGluLeuIleAlaAspLysIleSerProGlySerSerGlyLeuLysThr 858
             :::         ::::::   |||:::               :::|||      :::
Db     68974 GAAATTATGAATGTTTTGAAGGATCAG------------GCATCAGACTCAATGTCATTC 69021

Qy       859 SerGlnLeuLeuGluAspIleArgSerGlnSerSerValAspValSerLeuGlnAspIle 878
             ::::::|||::::::   |||         |||         |||         ||||||
Db     69022 AATGAGCTTATAAAACAAATCAATGAACACTCTCAAGATAGAGTTGAGTCTTCTGATATC 69081

Qy       879 LysAsnAlaLeuGlySerLeuGlnGlyGluGlyPheLeuThrValHisGlyAspIleVal 898
             :::   ||||||      ||||||   |||      :::   |||   |||:::   |||
Db     69082 CAGGAAGCCTTGTCAAGATTGCAACAAGAGGACAAGGTCATTGTCCTTGGCGAGGGTGTA 69141

Qy       899 LysArg 900
             :::|||
Db     69142 AGGAGA 69147

RESULT 13
MUSMCDC21
LOCUS       MUSMCDC21               3221 bp    mRNA    linear   ROD 01-AUG-2002
DEFINITION  Mus musculus mRNA for mcdc21 protein, complete cds.
ACCESSION   D26089
VERSION     D26089.1  GI:940405
KEYWORDS    .
SOURCE      Mus musculus (house mouse)
  ORGANISM  Mus musculus
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Euarchontoglires; Glires; Rodentia;
            Sciurognathi; Muridae; Murinae; Mus.
REFERENCE   1
  AUTHORS   Kimura,H., Takizawa,N., Nozaki,N. and Sugimoto,K.
  TITLE     Molecular cloning of cDNA encoding mouse Cdc21 and CDC46 homologs
            and characterization of the products: physical interaction between
            P1(MCM3) and CDC46 proteins
  JOURNAL   Nucleic Acids Res. 23 (12), 2097-2104 (1995)
   PUBMED   7610039
REFERENCE   2  (bases 1 to 3221)
  AUTHORS   Kimura,H.
  JOURNAL   Unpublished
REFERENCE   3  (bases 1 to 3221)
  AUTHORS   Kimura,H.
  TITLE     Direct Submission
  JOURNAL   Submitted (06-DEC-1993) Hiroshi Kimura, Hokkaido University,
            Research Center for Molecular Genetics; Kita 10, Nishi 8, Kita-ku,
            Sapporo, Hokkaido 060, Japan (Tel:81-11-706-3588,
            Fax:81-11-737-0536)
FEATURES             Location/Qualifiers
     source          1..3221
                     /organism="Mus musculus"
                     /mol_type="mRNA"
                     /db_xref="taxon:10090"
                     /clone="mCdc21-16-2"
                     /sex="female"
                     /cell_line="FM3A"
                     /cell_type="mammary carcinoma"
                     /clone_lib="lambda gt10 from T. Sado"
     gene            1..3221
                     /gene="mCdc21"
     CDS             127..2715
                     /gene="mCdc21"
                     /function="possibly involved in DNA replication"
                     /note="put. mouse homolog of yeast cdc21"
                     /codon_start=1
                     /product="mcdc21 protein"
                     /protein_id="BAA05082.1"
                     /db_xref="GI:940406"
                     /translation="MSSPASTPSRRSSRRGRVTPTQSLRSEESRSSPNRRRRGEDSST
                     GELLPMPTSPGADLQSPPAQNALFSSPPQMHSLAIPLDFDVSSPLTYGTPSSRVEGTP
                     RSGVRGTPVRQRPDLGSARKGLQVDLQSDGAAAEDIVPSEQSLGQKLVIWGTDVNVAT
                     CKENFQRFLQCFTDPLAKEEENVGIDITQPLYMQQLGEINITGEPFLNVNCEHIKSFS
                     KNLYRQLISYPQEVIPTFDMAVNEIFFDRYPDSILEHQIQVRPFNALKTKSMRNLNPE
                     DIDQLITISGMVIRTSQLIPEMQEAFFQCQVCAHTTRVEIDRGRIAEPCSCVHCHTTH
                     SMALIHNRSFFSDKQMIKLQESPEDMPAGQTPHTIVLFAHNDLVDKVQPGDRVNVTGI
                     YRAVPIRVNPRVSNVKSVYKTHIDVIHYRKTDAKRLHGLDEEAEQKLFSEKRVKLLKE
                     LSRKPDIYERLASALAPSIYEHEDIKKGILLQLFGGTRKDFSHTGRGKFRAEINILLC
                     GDPGTSKSQLLQYVYNLVPRGQYTSGKGSSAVGLTAYVMKDPETRQLVLQTGALVLSD
                     NGICCIDEFDKMNESTRSVLHEVMEQQTLSIAKAGIICQLNARTSVLAAANPIESQWN
                     PKKTTIENIQLPHTLLSRFDLIFLMLDPQDEAYDRRLAHHLVSLYYQSEEQVEEEFLD
                     MAVLKDYIAYAHSTIMPRLSEEASQALIEAYVNMRKIGSSRGMVSAYPRQLESLIRLA
                     EAHAKVRFSNKVEAIDVEEAKRLHREALKQSATDPRTGIVDISILTTGMSATSRKRKE
                     ELAEALRKLILSKGKTPALKYQQLFEDIRGQSDTAITKDMFEEALRALADDDFLTVTG
                     KTVRLL"
     polyA_site      3221
                     /gene="mCdc21"
```

```
ORIGIN

Alignment Scores:
Pred. No.:              2.99e-59      Length:         3221
Score:                  1822.50       Matches:        403
Percent Similarity:     59.57%        Conservative:   151
Best Local Similarity:  43.33%        Mismatches:     267
Query Match:            39.54%        Indels:         109
DB:                     9             Gaps:           17

US-10-768-511-8 (1-901) x MUSMCDC21 (1-3221)

Qy     14 SerProTyrProSerGlnSerGluGlyValSerThrProLeuProGlnValThrSerPro 33
Db    112 AGCAACCACAGCACCATGTCGTCCCCGGCATCCACCCCG-------------------- 150

Qy     34 SerPheAspAsnAlaAlaSerProValAlaGlyArgArgAlaValArgGlnThrProThr 53
Db    151 -----------------AGCCGCCGCAGCAGCCGACGCGGA---CGAGTCACCCCAACC 189

Qy     54 SerAlaValArgArgArgGlyArgGluThrAspSerAlaArgArgArgArgSerArgSer 73
Db    190 CAGTCCCTTCGAAGTGAGGAAAGCAGGTCGTCACCCAATCGGAGACGTAGAGGCGAAGAT 249

Qy     74 ArgSerLeuGly------------------------------------------------ 77
Db    250 TCTTCCACCGGAGAGCTACTGCCAATGCCCACCTCACCAGGAGCCGACCTGCAGAGCCCA 309

Qy     78 ---------AsnSerValTyrSerSerPro----------------------------- 84
Db    310 CCTGCACAGAATGCCTTGTTTTCCAGCCCTCCTCAGATGCATTCTTTAGCTATTCCTTTG 369

Qy     85 ---TyrAspAlaGlyThrProGlyThrProGlyThrProValAlaThrProValTyrAla 103
Db    370 GACTTTGATGTTAGTTCACCATTGACA-------------------------TATGGC 402

Qy    104 ThrProValGlyThrProMetGlyThrPro---SerPheHisArgGlyThrProGlnTyr 122
Db    403 ACTCCCAGCTCGCGAGTGGAAGGAACCCCAAGAAGTGGGGTGAGAGGCACACCT---GTA 459

Qy    123 LysGlnArgSerGluLeuGlySerGlnGlyLysProLeuHisArgArgArgArgSerGln 142
Db    460 AGGCAGAGGCCAGATCTGGGCTCAGCACGAAAGGGTTTGCAGGTGGATCTGCAGTCT--- 516

Qy    143 SerArgGluProGlyHisArgSerProSerArgGluProSerAlaAspGlyArgProSer 162
Db    517 --------------------------------------------GATGGCGCAGCAGCA 531

Qy    163 GluSerAlaGluProAspAspThrLeuGlyGlyGluTyrAlaTyrValTrpGlyThrAsn 182
Db    532 GAAGACATCGTACCAAGTGAACAGTCTCTAGGCCAAAAGCTTGTGATTTGGGGAACAGAT 591

Qy    183 ValAsnIleProAspValLeuArgAlaIleArgArgPheLeuHisAsnTyr--------- 199
Db    592 GTGAATGTGGCAACATGTAAAGAGAATTTCAGAGATTCCTTCAGTGTTTTACTGATCCT 651

Qy    200 ---------------ArgSerSerAlaHisAspLeuAsnSerLys---TyrIleGlnIle 213
Db    652 CTGGCCAAAGAAGAAGAAAATGTTGGCATAGATATTACTCAACCTTTGTACATGCAACAA 711

Qy    214 IleGluGluThrValGluArgGluGluAspThrLeuAsnIleAspMetSerAspIleTyr 233
Db    712 CTTGGAGAGATTAATATTACAGGAGAGCCATTTTTAAATGTGAACTGCGAACACATAAAA 771

Qy    234 AspHisAspProAspLeuTyrAlaLysIleValArgTyrProLeuAspIleIleProLeu 253
Db    772 TCATTTAGCAAAAATCTGTATAGACAGCTCATCTCCTACCCACAGGAGGTTATACCAACC 831

Qy    254 LeuAspThrGluCysGlnGluValAlaThrSerLeuLeuPro------ThrPheGluLys 271
Db    832 TTTGACATGGCTGTCAATGAGATCTTCTTTGACCGTTATCCTGACTCCATCTTAGAACAT 891

Qy    272 HisIleGluAlaArgProPheAsnLeuLysAlaSerValHisMetArgGluLeuAsnPro 291
Db    892 CAGATTCAAGTCAGACCTTTTAATGCGTTGAAGACAAAGAGTATGAGAAACTTGAATCCA 951

Qy    292 SerAspIleAspLysLeuValSerValLysGlyMetValIleArgCysSerSerIleIle 311
Db    952 GAAGACATTGATCAGCTCATCACCATCAGTGGCATGGTCATCAGAACATCACAGCTGATT 1011

Qy    312 ProGluIleLysGlyAlaPhePheLysCysLeuValCysGlyHisSerProProLeuVal 331
Db   1012 CCGGAGATGCAGGAGGCCTTTTTCCAATGCCAAGTCTGTGCCCACACC---ACCCGGGTG 1068

Qy    332 ThrValValLysGlyArgValGluGluProThrArgCysGluLysProGluCysAlaAla 351
Db   1069 GAGATAGATCGAGGCAGAATTGCTGAGCCCTGCAGTTGT------GTGCACTGCCACACT 1122

Qy    352 ArgAsnAlaMetSerLeuIleHisAsnArgCysThrPheAlaAsnLysGlnIleValArg 371
Db   1123 ACCCACAGCATGGCACTGATCCACAACCGATCATTCTTCTCTGACAAGCAAATGATCAAA 1182

Qy    372 LeuGlnGluThrProAspAlaIleProGluGlyGluThrProHisThrValSerMetCys 391
Db   1183 CTTCAAGAGTCTCCTGAAGACATGCCTGCTGGGCAGACACCTCACACTATTGTCCTTTTT 1242

Qy    392 LeuTyrAsnThrMetValAspAlaValLysProGlyAspArgIleGluValThrGlyVal 411
Db   1243 GCCCACAATGACCTTGTTGACAAGGTTCAACCAGGGGACAGAGTGAACGTCACAGGCATA 1302

Qy    412 PheLysAlaMetAlaValArgValGlyProAsnGlnArgThrLeuArgAlaLeuTyrLys 431
Db   1303 TATCGAGCAGTACCAATTCGAGTTAATCCAAGAGTGAGCAACGTGAAGTCTGTCTATAAA 1362

Qy    432 ThrTyrIleAspCysValHisValLysLysSerAspArgGlyArgLeuGlnThrGluAsp 451
Db   1363 ACCCACATTGATGTCATTCATTATCGGAAAACGGATGCAAAACGTCTGCAT--------- 1413

Qy    452 ProMetGluMetAspLysGluAsnAspMetTyrAlaGlyTyrHisGluSerAspThrSer 471
Db   1414 ----------------------------------GGCCTTGATGAAGAAGCAGAACAG 1437

Qy    472 GluAlaAlaAsnGluAlaLysIleGlnLysLeuLysGluLeuSerLysLeuProGlyIle 491
Db   1438 AAACTTTTTTCAGAGAAACGTGTGAAATTGCTTAAGGAACTTTCCAGGAAGCCAGATATT 1497

Qy    492 TyrAspArgLeuSerArgSerLeuAlaProSerIleTrpGluLeuGluAspIleLysLys 511
Db   1498 TATGAGCGGCTTGCTTCAGCCTTGGCTCCCAGCATTTATGAACATGAAGATATCAAAAAG 1557

Qy    512 GlyLeuLeuCysGlnLeuPheGlyGlyLysAlaLysLysIle------ProSerGlyAla 529
Db   1558 GGAATCTTACTTCAGCTCTTTGGTGGAACAAGGAAGGATTTCAGTCACACTGGAGGGGT 1617

Qy    530 SerPheArgGlyAspIleAsnValLeuLeuValGlyAspProGlyThrSerLysSerGln 549
Db   1618 AAATTCCGTGCTGAGATCAACATCCTTCTGTGTGGGGACCCTGGCACCAGCAAGTCCCAG 1677

Qy    550 LeuLeuGlnTyrValHisLysIleAlaProArgGlyIleTyrThrSerGlyArgGlySer 569
Db   1678 CTGCTACAGTATGTGTACAACCTGGTCCCCAGAGGCCAGTACACGTCTGGAAAAGGCTCC 1737

Qy    570 SerAlaValGlyLeuThrAlaTyrValThrLysAspProGluThrArgGluThrValLeu 589
Db   1738 AGTGCGGTCGGCCTCACCGCCTATGTGATGAAAGACCCTGAGACCAGGCAGCTTGTCCTC 1797

Qy    590 GluSerGlyAlaLeuValLeuSerAspArgGlyIleCysCysIleAspGluPheAspLys 609
Db   1798 CAGACAGGTGCCCTCGTCCTGAGTGACAATGGGATATGCTGCATCGATGAGTTTGACAAA 1857

Qy    610 MetSerAspAsnAlaArgSerMetLeuHisGluValMetGluGlnGlnThrValSerVal 629
Db   1858 ATGAATGAAAGCACAAGGTCTGTGCTGCATGAGGTCATGGAACAGCAGACTCTGTCCATT 1917

Qy    630 AlaLysGlyGlyIleIleAlaSerLeuAsnAlaArgThrSerValLeuAlaCysAlaAsn 649
```

```
Db       1918 GCAAAGGCTGGGATCATCTGTCAGCTCAATGCGCGCACCTCTGTCCTGGCAGCAGCAAAT 1977

Qy        650 ProSerGlySerArgTyrAsnAlaArgLeuSerValIleAspAsnIleGlnLeuProPro 669
              |||    |||:::::|||   :::   :::   |||:::||||||||||||||
Db       1978 CCTATTGAGTCTCAGTGGAATCCTAAAAAAACAACCATTGAAAATATCCAACTACCGCAC 2037

Qy        670 ThrLeuLeuSerArgPheAspLeuIleTyrLeuMetLeuAspLysProAspGluGlnAsn 689
              ||||||||||||||||||||||||||:::|||||||||||       ||||||
Db       2038 ACATTGTTGTCAAGGTTTGATCTCATTTTCCTCATGCTAGACCCTCAGGATGAGGCATAT 2097

Qy        690 AspArgArgLeuAlaArgHisLeuValAlaLeuHisTyrGluAsnTyrGluValSerLys 709
              ||||||||||||||   |||||||||:::|||:::|||::::::   |||      :::
Db       2098 GACCGGCGTCTAGCTCATCACCTGGTTTCATTGTACTACCAAAGTGAGGAGCAAGTGGAG 2157

Qy        710 GlnAspAlaLeuAspLeuGlnThrLeuThrAlaTyrIleThrTyrAlaArgGlnHisVal 729
              ::::::   |||||||:::         |||      ||||||   ||||||       :::
Db       2158 GAGGAGTTCCTGGACATGGCCGTGCTGAAAGACTACATTGCATATGCCCATAGTACCATC 2217

Qy        730 HisProThrLeuSerAspGluAlaAlaGluAspLeuIleAsnGlyTyrValGluMetArg 749
                 |||   ||||||:::||||||::::::    ||||||      ||||||    ||||||
Db       2218 ATGCCCCGACTGAGTGAGGAGGCCAGCCAGGCTCTCATTGAGGCTTATGTAAACATGAGG 2277

Qy        750 GlnLysGlyAsnPheProGlySerSerLysLysValIleThrAlaThrProArgGlnLeu 769
              :::                |||||||||:::   :::::::::|||   ||||||||||||
Db       2278 AAG------------ATTGGGAGTAGCCGGGGGATGGTTTCTGCTTACCCTCGACAGCTA 2325

Qy        770 GluSerMetIleArgIleSerGluAlaLeuAlaArgMetArgPheSerGluValValGlu 789
              ||||||:::||||||:::::::||||||      |||:::::::||||||||||      ||||||
Db       2326 GAGTCATTAATTCGCTTAGCAGAAGCCCATGCTAAAGTAAGATTTTCAAACAAAGTTGAA 2385

Qy        790 LysValAspAlaAlaGluAlaValArgLeuLeuAspValAlaLeuGlnGlnSerAlaThr 809
                 :::|||       ||||||   ||||||         ||||||:::||||||||||||
Db       2386 GCAATTGATGTGGAAGAGGCAAAACGCCTCCACCGGGAGGCTCTGAAGCAGTCTGCAACT 2445

Qy        810 AspHisAlaThrGlyThrIleAspMetAspLeuIleThrThrGlyValSerAlaSerGlu 829
              |||      ||||||     :::|||:::   ::::::::||||||:::||||||:::
Db       2446 GACCCTCGTACTGGCATTGTGGATATTTCTATTCTTACTACAGGAATGAGTGCCACTTCT 2505

Qy        830 ArgIleArgArgAlaAsnLeuLeuAlaAlaLeuArgGluLeuIleAlaAspLysIleSer 849
              |||   |||:::      |||         ||||||||:::||||||          :::|||
Db       2506 CGTAAACGGAAAGAAGAATTAGCTGAAGCATTGAGAAAACTTATT---------TTATCT 2556

Qy        850 ProGlySerSerSerGlyLeuLysThrSerGlnLeuLeuGluAspIleArgSerGlnSer 869
                 |||   :::      ||||||      ||||||   |||||||||||    ||||||
Db       2557 AAGGGTAAAACACCAGCCTTAAAGTACCAACAGCTGTTTGAGGATATTCGGGGACAGTCT 2616

Qy        870 SerValAspValSerLeuGlnAspIleLysAsnAlaLeuGlySerLeuGlnGlyGluGly 889
                       ::::::         :::   ||||||   :::|||       :::
Db       2617 GACACAGCAATTACCAAGGACATGTTTGAAGAAGCCCTGCGAGCTTTGGCTGATGATGAT 2676

Qy        890 PheLeuThrValHisGlyAspIleValLys 899
              ||||||||||||   |||      |||:::
Db       2677 TTCCTAACAGTGACTGGGAAGACTGTCCGC 2706
```

RESULT 14
AK220442

```
LOCUS       AK220442                3283 bp    mRNA    linear   ROD 01-MAR-2005
DEFINITION  Mus musculus mRNA for mKIAA4003 protein.
ACCESSION   AK220442
VERSION     AK220442.1  GI:60360103
KEYWORDS    FLI_CDNA.
SOURCE      Mus musculus (house mouse)
  ORGANISM  Mus musculus
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Euarchontoglires; Glires; Rodentia;
            Sciurognathi; Muridae; Murinae; Mus.
REFERENCE   1
  AUTHORS   Okazaki,N., Kikuno,R.F., Ohara,R., Inamoto,S., Nagase,T., Ohara,O.
            and Koga,H.
  TITLE     Prediction of the Coding Sequences of Mouse Homologues of KIAA
            Gene. The Complete Nucleotide Sequences of Mouse KIAA-homologous
            cDNAs Identified by Screening of Terminal sequences of cDNA Clones
            Randomly Sampled from Size-Fractionated Libraries.
  JOURNAL   Published Only in Database (2005)
REFERENCE   2  (bases 1 to 3283)
  AUTHORS   Okazaki,N., F.Kikuno,R., Nagase,T., Ohara,O. and Koga,H.
  TITLE     Direct Submission
  JOURNAL   Submitted (22-FEB-2005) Hisashi Koga, Kazusa DNA Research
            Institute, Laboratory for Genome Informatics; 2-6-7
            Kazusa-kamatari, Kisarazu, Chiba, 292-0818, Japan
            (E-mail:mouse@kazusa.or.jp, Tel:81-438-52-3919, Fax:81-438-52-3918)
FEATURES             Location/Qualifiers
     source          1..3283
                     /organism="Mus musculus"
                     /mol_type="mRNA"
                     /db_xref="taxon:10090"
                     /clone="mfj10260"
                     /tissue_type="fetal brain"
                     /note="vector:modified pBC SK+"
     gene            <1..3283
                     /gene="mKIAA4003"
     CDS             <678..2711
                     /gene="mKIAA4003"
                     /note="CDS is predicted by in silico analysis. Start codon
                     is not identified."
                     /codon_start=1
                     /evidence=not_experimental
                     /product="mKIAA4003 protein"
                     /protein_id="BAD90271.1"
                     /db_xref="GI:60360104"
                     /translation="ILLNLCTCNNLERLILQESHFLNVNCEHIKSFSKNLYRQLISYP
                     QEVIPTFDMAVNEIFFDRYPDSILEHQIQVRPFNALKTKSMRNLNPEDIDQLITISGM
                     VIRTSQLIPEMQEAFFQCQVCAHTTRVEIDRGRIAEPCSCVHCHTTHSMALIHNRSFF
                     SDKQMIKLQESPEDMPAGQTPHTIVLFAHNDLVDKVQPGDRVNVTGIYRAVPIRVNPR
                     VSNVKSVYKTHIDVIHYRKTDAKRLHGLDEEAEQKLFSEKRVKLLKELSRKPDIYERL
                     ASALAPSIYEHEDIKKGILLQLFGGTRKDFSHTGRGKFRAEINILLCGDPGTSKSQLL
                     QYVYNLVPRGQYTSGKGSSAVGLTAYVMKDPETRQLVLQTGALVLSDNGICCIDEFDK
                     MNESTRSVLHEVMEQQTLSIAKAGIICQLNARTSVLAAANPIESQWNPKKTTIENIQL
                     PHTLLSRFDLIFLMLDPQDEAYDRRLAHHLVSLYYQSEEQVEEEFLDMAVLKDYIAYA
                     HSTIMPRLSEEASQALIEAYVNMRKIGSSRGMVSAYPRQLESLIRLAEAHAKVRFSNK
                     VEAIDVEEAKRLHREALKQSATDPRTGIVDISILTTGMSATSRKRKEELAEALRKLIL
                     SKGKTPALKYQQLFEDIRGQSDTAITKDMFEEALRALADDDFLTVTGKTVRLL"
ORIGIN
```

```
Alignment Scores:
Pred. No.:               3.32e-59      Length:         3283
Score:                   1821.50       Matches:        403
Percent Similarity:      59.72%        Conservative:   153
Best Local Similarity:   43.29%        Mismatches:     265
Query Match:             39.52%        Indels:         110
DB:                      9             Gaps:           17
```

US-10-768-511-8 (1-901) x AK220442 (1-3283)

```
Qy         14 SerProTyrProSerGlnSerGluGlyValSerThrProLeuProGlnValThrSerPro 33
              |||   :::   :::   |||         |||||||||
Db        107 AGCAACCACAGCACCATGTCGTCCCCGGCATCCACCCCG-------------------- 145

Qy         34 SerPheAspAsnAlaAlaSerProValAlaGlyArgArgAlaValArgGlnThrProThr 53
                                 |||      :::   ||||||      |||   |||||||||
Db        146 ------------------AGCCGCCGCAGCAGCCGACGCGGA---CGAGTCACCCCAACC 184

Qy         54 SerAlaValArgArgArgGlyArgGluThrAspSerAlaArgArgArgArgSerArgSer 73
                 ::::::|||                 :::         ||||||||||||
Db        185 CAGTCCCTTCGAAGTGAGGAAAGCAGGTCGTCACCCAATCGGAGACGTAGAGGCGAAGAT 244

Qy         74 ArgSerLeuGly------------------------------------------------ 77
                 |||   |||
Db        245 TCTTCCACCGGAGAGCTACTGCCAATGCCCACCTCACCAGGAGCCGACCTGCAGAGCCCA 304

Qy         78 ---------AsnSerValTyrSerSerPro------------------------------ 84
                       |||::::::::|||||||||
Db        305 CCTGCACAGAATGCCTTGTTTCCAGCCCTCCTCAGATGCATTCTTTAGCTATTCCTTTG 364

Qy         85 ---TyrAspAlaGlyThrProGlyThrProGlyThrProValAlaThrProValTyrAla 103
                 :::|||      :::|||   |||                              |||
```

```
Db      365 GACTTTGATGTTAGTTCACCATTGACA--------------------------TATGGC 397

Qy      104 ThrProValGlyThrProMetGlyThrPro---SerPheHisArgGlyThrProGlnTyr 122

Db      398 ACTCCCAGCTCGCGAGTGGAAGGAACCCCAAGAAGTGGGTGAGAGGCACACCT---GTA 454

Qy      123 LysGlnArgSerGluLeuGlySerGlnGlyLysProLeuHisArgArgArgSerGln 142

Db      455 AGGCAGAGGCCAGATCTGGGCTCAGCACGAAAGGGTTTGCAGGTGGATCTGCAGTCT--- 511

Qy      143 SerArgGluProGlyHisArgSerProSerArgGluProSerAlaAspGlyArgProSer 162

Db      512 ---------------------------------------------GATGGCGCAGCAGCA 526

Qy      163 GluSerAlaGluProAspAspThrLeuGlyGlyGluTyrAlaTyrValTrpGlyThrAsn 182

Db      527 GAAGACATCGTACCAAGTGAACAGTCTCTAGGCCAAAAGCTTGTGATTTGGGGAACAGAT 586

Qy      183 ValAsnIleProAspValLeuArgAlaIleArgArgPheLeuHisAsnTyr--------- 199

Db      587 GTGAATGTGGCAACATGTAAAGAGAATTTTCAGAGATTCCTTCAGTGTTTTACTGATCCT 646

Qy      200 ---------------ArgSerSerAlaHisAspLeuAsnSerLys---TyrIleGln-Il 213

Db      647 CTGGCCAAAGAAGAAGAAAATGTTGGCATAGATATTACTCAACCTTTGTACATGCAACAA 706

Qy      213 eIleGluGluThrValGluArgGluGluAspThrLeuAsnIleAspMetSerAspIleTy 233

Db      707 CTTGGAGAGATTAATATTACAGGAGAGCCATTTTTTAAATGTGAACTGCGAACACATAAA 766

Qy      233 rAspHisAspProAspLeuTyrAlaLysIleValArgTyrProLeuAspIleIleProLe 253

Db      767 ATCATTTAGCAAAAATCTGTATAGACAGCTCATCTCCTACCCACAGGAGGTTATACCAAC 826

Qy      253 uLeuAspThrGluCysGlnGluValAlaThrSerLeuLeuPro------ThrPheGluLy 271

Db      827 CTTTGACATGGCTGTCAATGAGATCTTCTTTGACCGTTATCCTGACTCCATCTTAGAACA 886

Qy      271 sHisIleGluAlaArgProPheAsnLeuLysAlaSerValHisMetArgGluLeuAsnPr 291

Db      887 TCAGATTCAAGTCAGACCTTTTAATGCGTTGAAGACAAAGAGTATGAGAAACTTGAATCC 946

Qy      291 oSerAspIleAspLysLeuValSerValLysGlyMetValIleArgCysSerSerIleIl 311

Db      947 AGAAGACATTGATCAGCTCATCACCATCAGTGGCATGGTCATCAGAACATCACAGCTGAT 1006

Qy      311 eProGluIleLysGlyAlaPhePheLysCysLeuValCysGlyHisSerProProLeuVa 331

Db     1007 TCCGGAGATGCAGGAGGCCTTTTTCCAATGCCAAGTCTGTGCCCACACC---ACCCGGGT 1063

Qy      331 lThrValValLysGlyArgValGluGluProThrArgCysGluLysProGluCysAlaAl 351

Db     1064 GGAGATAGATCGAGGCAGAATTGCTGAGCCCTGCAGTTGT------GTGCACTGCCACAC 1117

Qy      351 aArgAsnAlaMetSerLeuIleHisAsnArgCysThrPheAlaAsnLysGlnIleValAr 371

Db     1118 TACCCACAGCATGGCACTGATCCACAACCGATCATTCTTCTCTGACAAGCAAATGATCAA 1177

Qy      371 gLeuGlnGluThrProAspAlaIleProGluGlyGluThrProHisThrValSerMetCy 391

Db     1178 ACTTCAAGAGTCTCCTGAAGACATGCCTGCTGGGCAGACACCTCACACTATTGTCCTTTT 1237

Qy      391 sLeuTyrAsnThrMetValAspAlaValLysProGlyAspArgIleGluValThrGlyVa 411

Db     1238 TGCCCACAATGACCTTGTTGACAAGGTTCAACCAGGGGACAGAGTGAACGTCACAGGCAT 1297

Qy      411 lPheLysAlaMetAlaValArgValGlyProAsnGlnArgThrLeuArgAlaLeuTyrLy 431

Db     1298 ATATCGAGCAGTACCAATTCGAGTTAATCCAAGAGTGAGCAACGTGAAGTCTGTCTATAA 1357

Qy      431 sThrTyrIleAspCysValHisValLysLysSerAspArgGlyArgLeuGlnThrGluAs 451

Db     1358 AACCCACATTGATGTCATTCATTATCGGAAAACGGATGCAAAACGTCTGCAT-------- 1409

Qy      451 pProMetGluMetAspLysGluAsnAspMetTyrAlaGlyTyrHisGluSerAspThrSe 471

Db     1410 ------------------------------------GGCCTTGATGAAGAAGCAGAACA 1432

Qy      471 rGluAlaAlaAsnGluAlaLysIleGlnLysLeuLysGluLeuSerLysLeuProGlyIl 491

Db     1433 GAAACTTTTTCAGAGAAACGTGTGAAATTGCTTAAGGAACTTTCCAGGAAGCCAGATAT 1492

Qy      491 eTyrAspArgLeuSerArgSerLeuAlaProSerIleTrpGluLeuGluAspIleLysLy 511

Db     1493 TTATGAGCGGCTTGCTTCAGCCTTGGCTCCCAGCATTTATGAACATGAAGATATCAAAAA 1552

Qy      511 sGlyLeuLeuCysGlnLeuPheGlyGlyLysAlaLysLysIle------ProSerGlyAl 529

Db     1553 GGGAATCTTACTTCAGCTCTTTGGTGGAACAAGGAAGGATTTCAGTCACACTGGGAGGGG 1612

Qy      529 aSerPheArgGlyAspIleAsnValLeuLeuValGlyAspProGlyThrSerLysSerGl 549

Db     1613 TAAATTCCGTGCTGAGATCAACATCCTTCTGTGTGGGGACCCTGGCACCAGCAAGTCCCA 1672

Qy      549 nLeuLeuGlnTyrValHisLysIleAlaProArgGlyIleTyrThrSerGlyArgGlySe 569

Db     1673 GCTGCTACAGTATGTGTACAACCTGGTCCCCAGAGGCCAGTACACGTCTGGAAAAGGCTC 1732

Qy      569 rSerAlaValGlyLeuThrAlaTyrValThrLysAspProGluThrArgGluThrValLe 589

Db     1733 CAGTGCGGTCGGCCTCACCGCCTATGTGATGAAAGACCCTGAGACCAGGCAGCTTGTCCT 1792

Qy      589 uGluSerGlyAlaLeuValLeuSerAspArgGlyIleCysCysIleAspGluPheAspLy 609

Db     1793 CCAGACAGGTGCCCTCGTCCTGAGTGACAATGGGATATGCTGCATCGATGAGTTTGACAA 1852

Qy      609 sMetSerAspAsnAlaArgSerMetLeuHisGluValMetGluGlnGlnThrValSerVa 629

Db     1853 AATGAATGAAAGCACAAGGTCTGTGCTGCATGAGGTCATGGAACAGCAGACTCTGTCCAT 1912

Qy      629 lAlaLysGlyGlyIleIleAlaSerLeuAsnAlaArgThrSerValLeuAlaCysAlaAs 649

Db     1913 TGCAAAGGCTGGGATCATCTGTCAGCTCAATGCGCGCACCTCTGTCCTGGCAGCAGCAAA 1972

Qy      649 nProSerGlySerArgTyrAsnAlaArgLeuSerValIleAspAsnIleGlnLeuProPr 669

Db     1973 TCCTATTGAGTCTCAGTGGAATCCTAAAAAAACAACCATTGAAAATATCCAACTACCGCA 2032

Qy      669 oThrLeuLeuSerArgPheAspLeuIleTyrLeuMetLeuAspLysProAspGluGlnAs 689

Db     2033 CACATTGTTGTCAAGGTTTGATCTCATTTTCCTCATGCTAGACCCTCAGGATGAGGCATA 2092

Qy      689 nAspArgArgLeuAlaArgHisLeuValAlaLeuHisTyrGluAsnTyrGluValSerLy 709

Db     2093 TGACCGGCGTCTAGCTCATCACCTGGTTTCATTGTACTACCAAAGTGAGGAGCAAGTGGA 2152

Qy      709 sGlnAspAlaLeuAspLeuGlnThrLeuThrAlaTyrIleThrTyrAlaArgGlnHisVa 729

Db     2153 GGAGGAGTTCCTGGACATGGCCGTGCTGAAAGACTACATTGCATATGCCCATAGTACCAT 2212

Qy      729 lHisProThrLeuSerAspGluAlaAlaGluAspLeuIleAsnGlyTyrValGluMetAr 749

Db     2213 CATGCCCCGACTGAGTGAGGAGGCCAGCCAGGCTCTCATTGAGGCTTATGTAAACATGAG 2272

Qy      749 gGlnLysGlyAsnPheProGlySerSerLysLysValIleThrAlaThrProArgGlnLe 769

Db     2273 GAAG------------ATTGGGAGTAGCCGGGGGATGGTTTCTGCTTACCCTCGACAGCT 2320

Qy      769 uGluSerMetIleArgIleSerGluAlaLeuAlaArgMetArgPheSerGluValValGl 789

Db     2321 AGAGTCATTAATTCGCTTAGCAGAAGCCCATGCTAAAGTAAGATTTTCAAACAAAGTTGA 2380

Qy      789 uLysValAspAlaAlaGluAlaValArgLeuLeuAspValAlaLeuGlnGlnSerAlaTh 809

Db     2381 AGCAATTGATGTGGAAGAGGCAAAACGCCTCCACCGGGAGGCTCTGAAGCAGTCTGCAAC 2440
```

```
Qy        809 rAspHisAlaThrGlyThrIleAspMetAspLeuIleThrThrGlyValSerAlaSerGl 829
              ||||     ||||||    :::|||:::   ::::::||||||||:::||||||:::
Db       2441 TGACCCTCGTACTGGCATTGTGGATATTTCTATTCTTACTACAGGAATGAGTGCCACTTC 2500

Qy        829 uArgIleArgArgAlaAsnLeuLeuAlaAlaLeuArgGluLeuIleAlaAspLysIleSe 849
              |||   |||:::          |||       ||||||||:::||||||          :::||
Db       2501 TCGTAAACGGAAAGAAGAATTAGCTGAAGCATTGAGAAAACTTATT---------TTATC 2551

Qy        849 rProGlySerSerSerGlyLeuLysThrSerGlnLeuLeuGluAspIleArgSerGlnSe 869
              |   |||   :::         ||||||      ||||||   |||||||||||   ||||||
Db       2552 TAAGGGTAAAACACCAGCCTTAAAGTACCAACAGCTGTTTGAGGATATTCGGGGACAGTC 2611

Qy        869 rSerValAspValSerLeuGlnAspIleLysAsnAlaLeuGlySerLeuGlnGlyGluGl 889
              |            ::::::           :::   ||||||   :::|||        :::
Db       2612 TGACACAGCAATTACCAAGGACATGTTTGAAGAAGCCCTGCGAGCTTTGGCTGATGATGA 2671

Qy        889 yPheLeuThrValHisGlyAspIleValLys 899
                ||||||||||||    |||      |||:::
Db       2672 TTTCCTAACAGTGACTGGGAAGACTGTCCGC 2702
```

RESULT 15
CR380957_02/c
WPCOMMENT
  Sequence split into 13 fragments LOCUS CR380957 Accession CR380957

Fragment Name	Begin	End
CR380957_00	1	110000
CR380957_01	100001	210000
CR380957_02	200001	310000
CR380957_03	300001	410000
CR380957_04	400001	510000
CR380957_05	500001	610000
CR380957_06	600001	710000
CR380957_07	700001	810000
CR380957_08	800001	910000
CR380957_09	900001	1010000
CR380957_10	1000001	1110000
CR380957_11	1100001	1210000
CR380957_12	1200001	1302002

Continuation (3 of 13) of CR380957 from base 200001 (CR380957 Candida glabrata strain CB

Alignment Scores:

Pred. No.:	1.27e-57	Length:	110000
Score:	1819.50	Matches:	397
Percent Similarity:	59.00%	Conservative:	170
Best Local Similarity:	41.31%	Mismatches:	261
Query Match:	39.48%	Indels:	133
DB:	15	Gaps:	21

US-10-768-511-8 (1-901) x CR380957_02 (1-110000)

```
Qy         31 ThrSerProSerPheAspAsnAlaAlaSerProValAlaGlyArgArgAlaValArgGln 50
              ||||||            :::|||   :::|||||||||   |||
Db      21810 ACTTCGGAGATGTCAAATAATCCCAGTTCTCCAGTGATGGGT------------------ 21769

Qy         51 ThrProThrSerAlaValArgArgArgGlyArgGluThrAspSerAlaArgArgArgArg 70
                 |||                           :::   |||:::         |||
Db      21768 ---CCTATT-----------------------GATGAGGGAAGTTCTTCTCCCGGCCGC 21736

Qy         71 SerArgSerArgSerLeuGly-----------------------------------Asn 78
              ::::::         |||:::|||                                |||
Db      21735 AACCAGCCCTCAAGCATGGGTCCAAATGGTGATCCATCCAGTCCAGCATTGTTTTTTAAC 21676

Qy         79 SerValTyrSerSerProTyrAspAlaGlyThrProGlyThrProGlyThrProValAla 98
              |||      |||   |||            |||         |||      :::|||:::
Db      21675 TCATCATCATCCCAACCACAGGGTGATTCTCAACCTGTCGCCCCATCTTCGCCAATTCAT 21616

Qy         99 ThrProValTyrAlaThrProValGlyThrProMetGlyThrProSerPheHisArgGly 118
                 |||         :::         ||||||   :::               :::|||
Db      21615 TTCCCA---------TCTTCCTCTAACAGACCAATGAGCTCAGATGTTCACTCTCAAGGA 21565

Qy        119 ThrProGlnTyrLysGlnArg-----------------------------------Ser 126
                       :::                                           |||
Db      21564 AGAAGGAGCAACAGAAATCACCTTGCAGCCAACAGATCGAGTGCTTTAGGAAGATTCTCT 21505

Qy        127 GluLeuGlySerGlnGlyLysProLeuHisArgArgArgSerGlnSerArgGluPro 146
              :::   ||||||      :::   ||||||            :::         |||
Db      21504 GACTACGGAAGCGATAGACAGGTACTACATAGTTCCTCTTCTAATATGGGCGGCTACCCT 21445

Qy        147 GlyHisArg---------------------------------------SerProSer 152
                 |||                                           ||||||
Db      21444 TCTCAGAGAACTAATTTAAGAAGAAATGATATTCACGCCTCAGATCTATCCTCACCAAGA 21385

Qy        153 ArgGluProSerAlaAspGlyArg------------------------------------ 160
              |||               |||   |||
Db      21384 AGGATTGTAGACTTCGACTCCAGGTCAGGTATACAGCAACCTTCTTCTTCTTCCTCTTCA 21325

Qy        161 ---ProSerGluSerAlaGluProAspAspThrLeuGlyGlyGluTyrAlaTyrValTrp 179
                 |||||||||:::   ||||||                              :::|||
Db      21324 ATGCCTTCCGAAGCTACCGAACCT---------------------CTAAGAATTATTTGG 21286

Qy        180 GlyThrAsnValAsnIleProAspValLeuArgAlaIleArgArgPheLeu--------- 196
              ||||||||||||:::|||      :::         :::   |||   ||||||
Db      21285 GGTACAAATGTCAGCATACAAGAATGTGCCAATAGCTTCAGGAATTTCCTGATGTCATTT 21226

Qy        197 ---HisAsnTyrArg---------SerSerAlaHisAspLeuAsnSerLys--------- 209
                 :::   ||||||                        |||:::
Db      21225 AAGTACAAGTATAGAAGAGTTTTGGATGGGAAGACAGATATAACCGATGATGAAGCAGAA 21166

Qy        210 ------------TyrIleGlnIleIleGluGluThrValGluArgGluGluAspThrLeu 225
                          |||::::::   :::   |||         |||               |||
Db      21165 GAAGAACTATACTATGTGAAGCAACTTAATGAAATGAGAGAACTTGGTACATCAAATTTA 21106

Qy        226 AsnIleAspMetSerAspIle------TyrAspHisAspProAspLeuTyrAlaLysIle 243
              |||:::|||         ::::::         :::         :::||||||   ::::::
Db      21105 AACCTCGATGCACGCAATTTGCTAGCATTTAAACAAACCGAGGAGTTGTATTATCAACTA 21046

Qy        244 ValArgTyrProLeuAspIleIleProLeuLeuAspThrGluCysGlnGluValAlaThr 263
              :::   ||||||   ::::::|||   ::::::|||            ::::::
Db      21045 TTGAACTATCCACAAGAAGTCATATCAATTATGGATCAAACCATCAAAGACTGTATGGTA 20986

Qy        264 SerLeu--------------------------LeuProThrPheGlu---LysHisIle 273
              ||||||                          |||         |||   |||
Db      20985 TCTCTTGTGGTTGATAACCAATTAGAACATGAGTTAGACGAAATTGAGTCGAAGTTTTAC 20926

Qy        274 GluAlaArgProPheAsnLeuLysAlaSerValHisMetArgGluLeuAsnProSerAsp 293
              :::   ||||||:::|||::::::                  |||||||||||||||:::|||
Db      20925 AAAGTAAGACCTTACAATGTTGAAACTCAAAAAGGTATGAGAGAATTGAATCCAAATGAT 20866

Qy        294 IleAspLysLeuValSerValLysGlyMetValIleArgCysSerSerIleIleProGlu 313
              ||||||||||||:::|||:::|||||||||:::|||:::|||   :::   :::||||||:::
Db      20865 ATAGATAAGTTGATAAGCCTGAAAGGGTTGGTTCTTAGAGCGACACCTGTTATACCAGAC 20806

Qy        314 IleLysGlyAlaPhePheLysCysLeuValCysGlyHisSerProProLeuValThrVal 333
              :::|||   |||||||||||||||       :::|||   |||:::         |||   :::
Db      20805 ATGAAAGTAGCTTTTTTCAAATGTAATATTTGTGATCACACC---ATGGCTGTTGAAATC 20749

Qy        334 ValLysGlyArgValGluGluProThrArgCysGluLysProGluCysAlaAlaArgAsn 353
                 :::|||   ::::::|||||   |||||||||:::   :::|||            |||
Db      20748 GATCGTGGTGTAATTCAGGAACCAGCTAGATGTGAACGTGTTGATTGTAATGAGGCTAAC 20689

Qy        354 AlaMetSerLeuIleHisAsnArgCysThrPheAlaAsnLysGlnIleValArgLeuGln 373
              :::|||:::||||||||||||||||||:::||||||:::|||||||:::::::::||||||
Db      20688 TCTATGACGCTAATTCATAATAGATGCTCTTTCGCTGATAAACAAGTTATTAAACTGCAA 20629

Qy        374 GluThrProAspAlaIleProGluGlyGluThrProHisThrValSerMetCysLeuTyr 393
              ||||||||||||      :::|||:::|||:::|||||||||:::||||||:::|||:::|||
Db      20628 GAAACTCCTGATTTGGTTCCAGATGGACAAACACCTCACTCTGTTTCTCTATGTGTTTAC 20569

Qy        394 AsnThrMetValAspAlaValLysProGlyAspArgIleGluValThrGlyValPheLys 413
              :::      :::||||||:::   :::         ||||||||||||||||||   |||:::
Db      20568 GACGAGTTAGTTGATAGTTGCCGTGCTGGTGACAGAATTGAAGTTACAGGTACCTTCAGA 20509

Qy        414 AlaMetAlaValArgValGlyProAsnGlnArgThrLeuArgAlaLeuTyrLysThrTyr 433
              ::::::   ::::::                  ||||||   |||::::::||||||||||||
Db      20508 TCAATCCCCATCAAGGCCAACTCTAGACAACGTGTCTTAAAATCTTTATACAAAACATAC 20449
```

```
Qy        434 IleAspCysValHisValLysLysSerAspArgGlyArgLeuGlnThrGluAspProMet 453
Db      20448 ATTGATGTCGTCCATGTAAAAAAGGTCTCGAACACTAGAATTGGTGTT---GATGTTTCA 20392

Qy        454 GluMetAspLysGlu-----------------------AsnAspMetTyrAlaGlyTyr 465
Db      20391 ACTATCGAACAAGAACTATTGCAAAATAAACTAGATAATAATGATGTC------------ 20344

Qy        466 HisGluSerAspThrSerGluAlaAlaAsnGluAlaLysIleGlnLysLeuLysGluLeu 485
Db      20343 ---------GAAGAAGTAAGACAAATCTCTGATGCTGAAATCGAGAAGATAAAGCAGGTT 20293

Qy        486 SerLysLeuProGlyIleTyrAspArgLeuSerArgSerLeuAlaProSerIleTrpGlu 505
Db      20292 GCACAAAGACCCGATCTTTATGATCTTCTTGCGAGATCTATAGCGCCAAGTATTTATGAA 20233

Qy        506 LeuGluAspIleLysLysGlyLeuLeuCysGlnLeuPheGlyGlyLysAlaLysLysIle 525
Db      20232 TTAGATGACGTCAAAAAGGGTATATTACTACAATTATTTGGTGGTACTAACAAAACGTTT 20173

Qy        526 ProSerGlyAlaSerPheArgGlyAspIleAsnValLeuLeuValGlyAspProGlyThr 545
Db      20172 AAGAAGGGTGGTCGTTATAGAGGTGATATTAACATTTTGCTATGTGGTGACCCCTCCACA 20113

Qy        546 SerLysSerGlnLeuLeuGlnTyrValHisLysIleAlaProArgGlyIleTyrThrSer 565
Db      20112 TCCAAATCCCAGATCTTACAATATGTTCATAAAATTGCTCCTAGAGGTGTCTACACATCT 20053

Qy        566 GlyArgGlySerSerAlaValGlyLeuThrAlaTyrValThrLysAspProGluThrArg 585
Db      20052 GGTAAGGGTTCATCAGCAGTAGGTCTAACGGCATATGTCACAAGAGATGTCGATAGTAAA 19993

Qy        586 GluThrValLeuGluSerGlyAlaLeuValLeuSerAspArgGlyIleCysCysIleAsp 605
Db      19992 CAATTAGTTTTAGAGAGTGGTGCCTTAGTGCTTTCTGATGGTGGTATCTGTTGTATTGAT 19933

Qy        606 GluPheAspLysMetSerAspAsnAlaArgSerMetLeuHisGluValMetGluGlnGln 625
Db      19932 GAATTCGATAAAATGAGTGAATCAACTAGATCTGTTCTTCACGAAGTGATGGAACAACAA 19873

Qy        626 ThrValSerValAlaLysGlyGlyIleIleAlaSerLeuAsnAlaArgThrSerValLeu 645
Db      19872 ACGATCTCCGTTGCTAAAGCTGGTATTATCACAACATTGAATGCTAGAAGTTCCATTTTA 19813

Qy        646 AlaCysAlaAsnProSerGlySerArgTyrAsnAlaArgLeuSerValIleAspAsnIle 665
Db      19812 GCTAGTGCAAACCCTATTGGATCCCGTTATAACCCTAATCTACCTGTTACTGAAAATATT 19753

Qy        666 GlnLeuProProThrLeuLeuSerArgPheAspLeuIleTyrLeuMetLeuAspLysPro 685
Db      19752 GATCTTCCACCCCCATTACTGTCAAGATTCGACCTTGTTTATATTATTTGGATAAAGTT 19693

Qy        686 AspGluGlnAsnAspArgArgLeuAlaArgHisLeuValAlaLeuHisTyrGluAsn--- 704
Db      19692 GATGAAAGCACTGACCGTGATCTGGCAAAACACTTGACCAGTCTTTATCTAGAGGACAAG 19633

Qy        705 ---TyrGluValSerLysGlnAspAlaLeuAspLeuGlnThrLeuThrAlaTyrIleThr 723
Db      19632 CCTGCACATGTTACTACCGATGATGTTCTCCCAATCGATTTCTTAACACAATACATCAAT 19573

Qy        724 TyrAlaArgGlnHisValHisProThrLeuSerAspGluAlaAlaGluAspLeuIleAsn 743
Db      19572 TATGTCAAACAGAATGTTCATCCACTTGTAACAGAGCAAGCAAAAAATGAACTTGTAAAG 19513

Qy        744 GlyTyrValGluMetArgGlnLysGlyAsnPheProGlySerSerLysLysValIleThr 763
Db      19512 GCTTATGTTGGCATGAGAAAAATGGGTGACGATTCAAGATCAGATGAAAAGAGAATTACT 19453

Qy        764 AlaThrProArgGlnLeuGluSerMetIleArgIleSerGluAlaLeuAlaArgMetArg 783
Db      19452 GCAACAACGAGACAGCTTGAAAGTATGATTCGTTTATCTGAAGCCCATGCCAAGATGAGA 19393

Qy        784 PheSerGluValValGluLysValAspAlaAlaGluAlaValArgLeuLeuAspValAla 803
Db      19392 TTATCAAGCACTGTAGATCTTGAAGACGTCCGTGAGGCTGTCAGATTAATGAAATCGGCT 19333

Qy        804 LeuGlnGlnSerAlaThrAspHisAlaThrGlyThrIleAspMetAspLeuIleThrThr 823
Db      19332 ATTAAAGATTACGCAACTGATCCAAAGACAGGTAAAATCGATATGAATTTGGTTCAAACT 19273

Qy        824 GlyValSerAlaSerGluArgIleArgArgAlaAsnLeu------------LeuAlaAla 839
Db      19272 GGTAAATCAGTTATACAGCGTAAATTACAAGAGGATTTGGCTAGGGAGATAATACGTATT 19213

Qy        840 LeuArgGluLeuIleAlaAspLysIleSerProGlySerSerSerGlyLeuLysThrSer 859
Db      19212 CTGAAGGAGTATCCAGCTGACTCTATGAGT------------------------TTCAAT 19177

Qy        860 GlnLeuLeuGluAspIleArgSerGlnSerSerValAspValSerLeuGlnAspIleLys 879
Db      19176 GAGTTGATCAAGCAAATCAATGAACAAGCACAAGACAGAGTAGAACCAAGTCAAGTGTCA 19117

Qy        880 AsnAlaLeuGlySerLeuGlnGlyGluGlyPheLeuThrValHisGlyAspIleValLys 899
Db      19116 AGCACTTTATCCAGGTTACAGCAGGAAGATAAAGTGATTATCCTGGGTGAAGGTGTTAGA 19057

Qy        900 Arg 900
Db      19056 AGA 19054
```

Search completed: December 6, 2005, 12:36:37
Job time : 8562 secs

```
OM protein - nucleic search, using frame_plus_p2n model

Run on:          December  6, 2005, 06:12:03 ; Search time 867 Seconds
                                               (without alignments)
                                               6926.053 Million cell updates/sec

Title:           US-10-768-511-8
Perfect score:   4609
Sequence:        1 MENNDALDIGAVSSPYPSQS..........LGSLQGEGFLTVHGDIVKRV 901

Scoring table:   BLOSUM62
                 Xgapop 10.0 , Xgapext  0.5
                 Ygapop 10.0 , Ygapext  0.5
                 Fgapop  6.0 , Fgapext  7.0
                 Delop   6.0 , Delext   7.0

Searched:        4996997 seqs, 3332346308 residues

Total number of hits satisfying chosen parameters:      9993994

Minimum DB seq length: 0
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%
                 Maximum Match 100%
                 Listing first 45 summaries

Command line parameters:
-MODEL=frame+_p2n.model -DEV=xlh
-Q=/cgn2_1/USPTO_spool/US10768511/runat_02122005_102333_18396/app_query.fasta_1.1095
-DB=N_Geneseq -QFMT=fastap -SUFFIX=rng -MINMATCH=0.1 -LOOPCL=0 -LOOPEXT=0
-UNITS=bits -START=1 -END=-1 -MATRIX=blosum62 -TRANS=human40.cdi -LIST=45
-DOCALIGN=200 -THR_SCORE=pct -THR_MAX=100 -THR_MIN=0 -ALIGN=15 -MODE=LOCAL
-OUTFMT=pto -NORM=ext -HEAPSIZE=500 -MINLEN=0 -MAXLEN=2000000000
-USER=US10768511_@CGN_1_1_608_@runat_02122005_102333_18396 -NCPU=6 -ICPU=3
-NO_MMAP -LARGEQUERY -NEG_SCORES=0 -WAIT -DSPBLOCK=100 -LONGLOG
-DEV_TIMEOUT=120 -WARN_TIMEOUT=30 -THREADS=1 -XGAPOP=10 -XGAPEXT=0.5 -FGAPOP=6
-FGAPEXT=7 -YGAPOP=10 -YGAPEXT=0.5 -DELOP=6 -DELEXT=7

Database :       N_Geneseq_21:*
                 1:  geneseqn1980s:*
                 2:  geneseqn1990s:*
                 3:  geneseqn2000s:*
                 4:  geneseqn2001as:*
                 5:  geneseqn2001bs:*
                 6:  geneseqn2002as:*
                 7:  geneseqn2002bs:*
                 8:  geneseqn2003as:*
                 9:  geneseqn2003bs:*
                 10:  geneseqn2003cs:*
                 11:  geneseqn2003ds:*
                 12:  geneseqn2004as:*
                 13:  geneseqn2004bs:*
                 14:  geneseqn2005s:*
```

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	% Query Match	Length	DB	ID	Description
1	4609	100.0	4348	6	AAI67613	Aai67613 Cell cycl
2	4609	100.0	4348	10	AAD60813	Aad60813 Physcomit
3	2550	55.3	2841	13	ADX53995	Adx53995 Plant ful
4	1883	40.9	3072	8	ABT21208	Abt21208 Aspergill
5	1871.5	40.6	2625	8	ABT19388	Abt19388 Aspergill
6	1840	39.9	1723	12	ADJ39580	Adj39580 Plant cDN
7	1835.5	39.8	2970	10	ADB69927	Adb69927 C. neofor
8	1833	39.8	2802	13	ADT47760	Adt47760 Bacterial
9	1819	39.5	2881	13	ADS49058	Ads49058 Bacterial
10	1817.5	39.4	3273	6	ABL65258	Abl65258 Lung canc
11	1817.5	39.4	3273	6	ABL65666	Abl65666 Lung canc
12	1817.5	39.4	3273	6	ABL65259	Abl65259 Lung canc
13	1817.5	39.4	3273	12	ADO19849	Ado19849 Human PRO
14	1817.5	39.4	3273	12	ADO19847	Ado19847 Human PRO
15	1817.5	39.4	3273	13	ADR26059	Adr26059 Breast ca
16	1817.5	39.4	3273	13	ADU05807	Adu05807 Novel bro
17	1816.5	39.4	3395	10	ADJ56481	Adj56481 Frog cDNA
18	1812	39.3	3248	13	ACN38801	Acn38801 Tumour-as
19	1811.5	39.3	2860	4	AAH17708	Aah17708 Human cDN
20	1791	38.9	3394	6	AAS94968	Aas94968 Human DNA
21	1789.5	38.8	2842	3	AAF15678	Aaf15678 Human pro
22	1764	38.3	3362	8	ABT20610	Abt20610 Aspergill
23	1764	38.3	4944	8	ABT18200	Abt18200 Aspergill
24	1764	38.3	5362	8	ABT20014	Abt20014 Aspergill
25	1752.5	38.0	2944	8	ABT18794	Abt18794 Aspergill
26	1752	38.0	2733	6	ABZ32193	Abz32193 Candida a
27	1733.5	37.6	5265	12	ADP83382	Adp83382 Breast sp
28	1690.5	36.7	2692	4	ABL02965	Abl02965 Drosophil
29	1677	36.4	5419	10	ADB69205	Adb69205 C. neofor
30	1675	36.3	3419	10	ADB69566	Adb69566 C. neofor
31	1670.5	36.2	1571	13	ADX36902	Adx36902 Plant ful
32	1634.5	35.5	3178	13	ACN41452	Acn41452 Human dia
33	1584.5	34.4	5065	4	ABL02964	Abl02964 Drosophil
34	1139	24.7	723	6	AAI67610	Aai67610 Cell cycl
35	1139	24.7	723	10	AAD60810	Aad60810 Physcomit
36	1070	23.2	2046	4	AAF57040	Aaf57040 DNA seque
37	1021.5	22.2	2001	13	ADT46787	Adt46787 Bacterial
38	1016.5	22.1	2280	13	ADT47813	Adt47813 Bacterial
39	1016	22.0	2821	12	ADQ09277	Adq09277 Human MCM
40	1003	21.8	2445	8	ABT20750	Abt20750 Aspergill
41	998	21.7	2427	6	ABL62640	Abl62640 Colon ade
42	998	21.7	2427	6	ABN95188	Abn95188 Gene #168
43	998	21.7	2427	12	ADO19789	Ado19789 Human PRO
44	998	21.7	2427	12	ADO19648	Ado19648 Human PRO
45	998	21.7	2427	13	ADR24507	Adr24507 Breast ca

ALIGNMENTS

```
RESULT 1
AAI67613
ID   AAI67613 standard; cDNA; 4348 BP.
XX
AC   AAI67613;
XX
DT   27-FEB-2002  (first entry)
XX
DE   Cell cycle protein 2 (CC-2) encoding cDNA.
XX
KW   Cell Cycle Stress-Related Protein; CCSRP; cell cycle protein; CC-1; CC-2;
KW   CC-3; environmental stress; ss.
XX
OS   Physcomitrella patens.
XX
PN   WO200177354-A2.
XX
PD   18-OCT-2001.
XX
PF   06-APR-2001; 2001WO-US011294.
XX
PR   07-APR-2000; 2000US-0196001P.
XX
PA   (BADI ) BASF PLANT SCI GMBH.
XX
PI   Costa E SilvaO,  Bohnert HJ,  Van Thielen N,  Chen R;
PI   Sarria-Millan R;
XX
```

```
DR   WPI; 2002-049151/06.
DR   P-PSDB; AAG66003.
XX
PT   Novel Cell Cycle Stress-Related Protein useful for increasing tolerance
PT   to environmental stress, is selected from Cell Cycle Proteins 1-3, or
PT   their orthologs.
XX
PS   Claim 4; Fig 2B; 90pp; English.
XX
CC   The invention relates to a Cell Cycle Stress-Related Protein (CCSRP),
CC   isolated from Physcomitrella patens, and selected from Cell Cycle (CC)-1
CC   protein, CC-2 protein, CC-3 protein, or their orthologs. The CCSRP and
CC   encoding nucleic acids are useful for increasing tolerance to
CC   environmental stress selected from salinity, drought and temperature, in
CC   transgenic plants including monocot and dicot selected from maize, wheat,
CC   rye, oat, triticale, rice, barley, soybean, peanut, cotton, rapeseed,
CC   canola, manihot, pepper, sunflower, tagetes, solanaceous plants, potato,
CC   tobacco, eggplant, tomato, Vicia species, pea, alfalfa, coffee, cacao,
CC   tea, Salix species, oil palm, coconut, perennial grass and forage crops.
CC   The nucleic acid is also useful for identifying organisms e.g. P. patens
CC   in a mixed population of microorganisms. The nucleic acids are also
CC   useful for evolutionary and protein structural studies. The proteins and
CC   nucleic acids are useful as markers for specific regions of the genome.
CC   The present sequence represents the cDNA encoding P. patens CC-2 protein
XX
SQ   Sequence 4348 BP; 1120 A; 987 C; 1093 G; 1148 T; 0 U; 0 Other;

Alignment Scores:
Pred. No.:              9.31e-263      Length:         4348
Score:                  4609.00        Matches:        901
Percent Similarity:     100.00%        Conservative:   0
Best Local Similarity:  100.00%        Mismatches:     0
Query Match:            100.00%        Indels:         0
DB:                     6              Gaps:           0

US-10-768-511-8 (1-901) x AAI67613 (1-4348)

Qy          1 MetGluAsnAsnAspAlaLeuAspIleGlyAlaValSerSerProTyrProSerGlnSer 20
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1554 ATGGAAAATAATGATGCACTTGACATTGGAGCCGTGTCGTCCCCATATCCTTCGCAATCT 1613

Qy         21 GluGlyValSerThrProLeuProGlnValThrSerProSerPheAspAsnAlaAlaSer 40
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1614 GAAGGAGTGTCTACGCCATTGCCGCAAGTAACATCACCGAGCTTCGACAATGCAGCCTCA 1673

Qy         41 ProValAlaGlyArgArgAlaValArgGlnThrProThrSerAlaValArgArgArgGly 60
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1674 CCCGTGGCCGGGCGGAGGGCCGTACGGCAGACCCCTACATCTGCAGTTCGAAGGAGAGGG 1733

Qy         61 ArgGluThrAspSerAlaArgArgArgArgSerArgSerArgSerLeuGlyAsnSerVal 80
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1734 AGAGAAACGGATTCCGCTCGTCGTAGGAGGAGTCGATCTCGCAGTTTAGGCAATTCTGTT 1793

Qy         81 TyrSerSerProTyrAspAlaGlyThrProGlyThrProGlyThrProValAlaThrPro 100
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1794 TATAGTTCCCCTTACGATGCGGGGACTCCTGGAACTCCTGGAACTCCAGTGGCTACTCCG 1853

Qy        101 ValTyrAlaThrProValGlyThrProMetGlyThrProSerPheHisArgGlyThrPro 120
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1854 GTTTACGCTACCCCAGTCGGTACACCTATGGGTACCCCATCGTTCCATCGTGGCACGCCA 1913

Qy        121 GlnTyrLysGlnArgSerGluLeuGlySerGlnGlyLysProLeuHisArgArgArgArg 140
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1914 CAGTACAAACAGCGCAGTGAGCTTGGTTCCCAGGGGAAGCCTCTACATCGGAGACGTCGA 1973

Qy        141 SerGlnSerArgGluProGlyHisArgSerProSerArgGluProSerAlaAspGlyArg 160
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1974 TCTCAATCCAGAGAACCCGGGCATCGATCTCCTTCAAGGGAACCTAGTGCTGATGGGCGT 2033

Qy        161 ProSerGluSerAlaGluProAspAspThrLeuGlyGlyGluTyrAlaTyrValTrpGly 180
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       2034 CCCTCTGAATCTGCTGAGCCAGATGACACTTTGGGTGGAGAATATGCTTATGTTTGGGGG 2093

Qy        181 ThrAsnValAsnIleProAspValLeuArgAlaIleArgArgPheLeuHisAsnTyrArg 200
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       2094 ACGAATGTTAACATTCCAGATGTGCTTAGGGCGATTCGTCGATTTCTCCACAATTATCGT 2153

Qy        201 SerSerAlaHisAspLeuAsnSerLysTyrIleGlnIleIleGluGluThrValGluArg 220
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       2154 TCGAGTGCTCATGATCTTAATTCCAAGTACATCCAGATCATAGAGGAGACTGTGGAGCGT 2213

Qy        221 GluGluAspThrLeuAsnIleAspMetSerAspIleTyrAspHisAspProAspLeuTyr 240
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       2214 GAGGAGGATACTCTAAATATCGACATGTCAGACATTTATGACCATGATCCTGATCTATAC 2273

Qy        241 AlaLysIleValArgTyrProLeuAspIleIleProLeuLeuAspThrGluCysGlnGlu 260
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       2274 GCAAAAATTGTTCGATACCCACTCGACATCATCCCCCTGTTGGACACTGAGTGTCAGGAA 2333

Qy        261 ValAlaThrSerLeuLeuProThrPheGluLysHisIleGluAlaArgProPheAsnLeu 280
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       2334 GTTGCTACCTCTTTACTACCAACGTTTGAGAAGCATATTGAGGCCAGACCTTTCAATCTC 2393

Qy        281 LysAlaSerValHisMetArgGluLeuAsnProSerAspIleAspLysLeuValSerVal 300
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       2394 AAAGCATCGGTGCACATGCGTGAACTCAACCCTTCAGATATAGACAAATTGGTTTCTGTT 2453

Qy        301 LysGlyMetValIleArgCysSerSerIleIleProGluIleLysGlyAlaPhePheLys 320
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       2454 AAAGGAATGGTTATCCGGTGCAGTTCTATCATACCTGAAATTAAGGGGGCCTTCTTCAAA 2513

Qy        321 CysLeuValCysGlyHisSerProProLeuValThrValValLysGlyArgValGluGlu 340
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       2514 TGTTTAGTGTGTGGTCACTCGCCTCCGCTAGTTACAGTTGTTAAAGGGCGGGTTGAGGAG 2573

Qy        341 ProThrArgCysGluLysProGluCysAlaAlaArgAsnAlaMetSerLeuIleHisAsn 360
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       2574 CCAACAAGGTGTGAAAAGCCAGAATGTGCAGCACGGAATGCTATGTCTCTTATTCACAAT 2633

Qy        361 ArgCysThrPheAlaAsnLysGlnIleValArgLeuGlnGluThrProAspAlaIlePro 380
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       2634 CGATGCACTTTTGCAAATAAGCAGATAGTGCGTCTTCAAGAAACTCCAGATGCCATTCCT 2693

Qy        381 GluGlyGluThrProHisThrValSerMetCysLeuTyrAsnThrMetValAspAlaVal 400
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       2694 GAAGGAGAGACTCCACACACAGTCAGCATGTGTTTATACAACACTATGGTTGATGCTGTG 2753

Qy        401 LysProGlyAspArgIleGluValThrGlyValPheLysAlaMetAlaValArgValGly 420
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       2754 AAGCCTGGAGATCGTATTGAGGTAACAGGAGTTTTCAAGGCCATGGCAGTTCGAGTTGGT 2813

Qy        421 ProAsnGlnArgThrLeuArgAlaLeuTyrLysThrTyrIleAspCysValHisValLys 440
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       2814 CCGAATCAACGAACATTACGAGCATTGTATAAGACCTACATCGATTGCGTGCACGTCAAG 2873

Qy        441 LysSerAspArgGlyArgLeuGlnThrGluAspProMetGluMetAspLysGluAsnAsp 460
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       2874 AAGTCTGACAGGGGTCGACTGCAAACTGAAGATCCTATGGAGATGGATAAGGAGAATGAT 2933

Qy        461 MetTyrAlaGlyTyrHisGluSerAspThrSerGluAlaAlaAsnGluAlaLysIleGln 480
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       2934 ATGTATGCTGGGTATCATGAAAGTGATACTTCAGAAGCTGCTAATGAAGCAAAGATTCAA 2993

Qy        481 LysLeuLysGluLeuSerLysLeuProGlyIleTyrAspArgLeuSerArgSerLeuAla 500
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       2994 AAACTTAAAGAGCTGTCCAAGCTCCCGGGCATTTATGATAGACTTTCAAGGTCGCTGGCT 3053

Qy        501 ProSerIleTrpGluLeuGluAspIleLysLysGlyLeuLeuCysGlnLeuPheGlyGly 520
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       3054 CCAAGCATTTGGGAGCTTGAAGATATTAAAAAGGGTCTTCTTTGCCAGCTCTTTGGTGGG 3113

Qy        521 LysAlaLysLysIleProSerGlyAlaSerPheArgGlyAspIleAsnValLeuLeuVal 540
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       3114 AAGGCTAAGAAAATTCCATCTGGAGCATCTTTCCGAGGTGACATCAATGTTTTACTTGTT 3173

Qy        541 GlyAspProGlyThrSerLysSerGlnLeuLeuGlnTyrValHisLysIleAlaProArg 560
```

```
                |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       3174 GGGGACCCTGGTACCAGTAAATCTCAGCTGCTTCAGTATGTGCACAAGATAGCTCCTCGT 3233

Qy        561 GlyIleTyrThrSerGlyArgGlySerSerAlaValGlyLeuThrAlaTyrValThrLys 580
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       3234 GGAATCTACACTAGTGGGCGAGGAAGTTCGGCGGTTGGGCTGACAGCGTATGTAACGAAG 3293

Qy        581 AspProGluThrArgGluThrValLeuGluSerGlyAlaLeuValLeuSerAspArgGly 600
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       3294 GATCCAGAAACTCGAGAGACGGTATTGGAGAGCGGAGCTTTGGTTCTTAGTGATCGTGGG 3353

Qy        601 IleCysCysIleAspGluPheAspLysMetSerAspAsnAlaArgSerMetLeuHisGlu 620
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       3354 ATATGCTGTATCGATGAGTTCGACAAAATGTCTGATAATGCCCGAAGCATGCTTCATGAG 3413

Qy        621 ValMetGluGlnGlnThrValSerValAlaLysGlyGlyIleIleAlaSerLeuAsnAla 640
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       3414 GTAATGGAGCAACAAACGGTATCTGTAGCCAAAGGGGGTATCATTGCCTCGCTGAACGCT 3473

Qy        641 ArgThrSerValLeuAlaCysAlaAsnProSerGlySerArgTyrAsnAlaArgLeuSer 660
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       3474 CGGACGTCTGTCCTTGCATGTGCAAATCCTAGTGGGTCCCGATACAATGCGCGCCTTTCT 3533

Qy        661 ValIleAspAsnIleGlnLeuProProThrLeuLeuSerArgPheAspLeuIleTyrLeu 680
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       3534 GTGATTGATAACATCCAGCTTCCTCCAACTCTACTTTCTAGATTTGATTTAATTTACTTA 3593

Qy        681 MetLeuAspLysProAspGluGlnAsnAspArgArgLeuAlaArgHisLeuValAlaLeu 700
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       3594 ATGCTCGACAAACCAGACGAGCAAAACGATCGTCGTCTCGCCAGGCATCTCGTGGCTTTA 3653

Qy        701 HisTyrGluAsnTyrGluValSerLysGlnAspAlaLeuAspLeuGlnThrLeuThrAla 720
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       3654 CACTATGAAAACTATGAAGTTTCAAAGCAGGACGCCTTAGATCTACAAACACTTACCGCG 3713

Qy        721 TyrIleThrTyrAlaArgGlnHisValHisProThrLeuSerAspGluAlaAlaGluAsp 740
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       3714 TATATCACCTATGCTCGTCAGCATGTACATCCTACATTAAGTGATGAAGCTGCTGAAGAT 3773

Qy        741 LeuIleAsnGlyTyrValGluMetArgGlnLysGlyAsnPheProGlySerSerLysLys 760
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       3774 TTGATTAATGGCTATGTTGAGATGCGCCAAAAGGGCAACTTTCCTGGAAGCAGTAAAAAG 3833

Qy        761 ValIleThrAlaThrProArgGlnLeuGluSerMetIleArgIleSerGluAlaLeuAla 780
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       3834 GTGATAACAGCCACACCTCGGCAACTCGAAAGTATGATTCGTATCAGTGAAGCCCTAGCT 3893

Qy        781 ArgMetArgPheSerGluValValGluLysValAspAlaAlaGluAlaValArgLeuLeu 800
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       3894 CGAATGAGATTTTCTGAAGTGGTAGAGAAAGTTGATGCAGCAGAAGCTGTGCGCCTTTTA 3953

Qy        801 AspValAlaLeuGlnGlnSerAlaThrAspHisAlaThrGlyThrIleAspMetAspLeu 820
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       3954 GACGTCGCTTTGCAGCAATCTGCTACTGATCATGCAACAGGTACGATAGACATGGATCTT 4013

Qy        821 IleThrThrGlyValSerAlaSerGluArgIleArgArgAlaAsnLeuLeuAlaAlaLeu 840
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       4014 ATCACGACTGGAGTGTCGGCCAGCGAGCGTATTCGTCGGGCCAACTTGCTAGCTGCTCTG 4073

Qy        841 ArgGluLeuIleAlaAspLysIleSerProGlySerSerSerGlyLeuLysThrSerGln 860
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       4074 CGAGAGCTTATAGCAGATAAAATTTCACCTGGCAGCTCCTCTGGCTTGAAGACCAGTCAG 4133

Qy        861 LeuLeuGluAspIleArgSerGlnSerSerValAspValSerLeuGlnAspIleLysAsn 880
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       4134 CTTCTTGAGGATATCCGGAGCCAAAGCAGTGTGGACGTTAGTTTGCAGGATATTAAAAAT 4193

Qy        881 AlaLeuGlySerLeuGlnGlyGluGlyPheLeuThrValHisGlyAspIleValLysArg 900
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       4194 GCTCTGGGTAGCCTCCAAGGAGAAGGCTTTCTTACTGTCCATGGTGACATAGTCAAGAGA 4253

Qy        901 Val 901
              |||
Db       4254 GTT 4256

RESULT 2
AAD60813
ID   AAD60813 standard; cDNA; 4348 BP.
XX
AC   AAD60813;
XX
DT   15-JAN-2004  (first entry)
XX
DE   Physcomitrella patens cell cycle protein (CC-2) cDNA.
XX
KW   Cell cycle stress-related protein; CCSRP; metal; cell cycle protein; CC;
KW   transgenic plant; environmental stress; stress tolerance; salinity;
KW   drought; temperature; chemical; oxidative stress; gene; ss.
XX
OS   Physcomitrella patens.
XX
FH   Key             Location/Qualifiers
FT   CDS             147. .4259
FT                   /*tag= a
FT                   /product= "CC-2 protein"
FT                   /transl_except= (pos:1404. .2816, aa:Gly-Pro)
XX
PN   US2003097675-A1.
XX
PD   22-MAY-2003.
XX
PF   06-APR-2001; 2001US-00828062.
XX
PR   07-APR-2000; 2000US-0196001P.
XX
PA   (SILV/) COSTA E SILVA O.
PA   (BOHN/) BOHNERT H J.
PA   (THIE/) THIELEN N V.
PA   (CHEN/) CHEN R.
PA   (SARR/) SARRIA-MILLAN R.
XX
PI   Costa E SilvaO,  Bohnert HJ,  Thielen NV,  Chen R,  Sarria-Millan R;
XX
DR   WPI; 2003-765533/72.
DR   P-PSDB; AAE39980.
XX
PT   Novel cell cycle stress-related protein capable of conferring stress
PT   tolerance such as tolerance towards salinity, drought, temperature,
PT   chemical, pathogens, to plants upon over-expression.
XX
PS   Claim 16; Fig 2B; 62pp; English.
XX
CC   The invention relates to an isolated cell cycle stress-related protein
CC   (CCSRP) which is chosen from a cell cycle-1 protein (CC-1), CC-2, CC-3
CC   and its orthologues. Recombinant expression vector is useful for
CC   producing a transgenic plant containing a CCSRP coding nucleic acid,
CC   where expression of the nucleic acid in the plant results in increased
CC   tolerance to environmental stress as compared to a wild type variety of
CC   the plant which involves transforming a plant cell with the recombinant
CC   expression vector, generating from the plant cell a transgenic plant with
CC   an increased tolerance to environmental stress as compared to a wild type
CC   variety of the plant. CCSRP is useful for conferring stress tolerance
CC   such as tolerance towards salinity, drought, temperature, metal,
CC   chemical, pathogens and oxidative stress or their combinations to plants.
CC   The present sequence is Physcomitrella patens CC-2 cDNA
XX
SQ   Sequence 4348 BP; 1120 A; 987 C; 1093 G; 1148 T; 0 U; 0 Other;

Alignment Scores:
Pred. No.:              9.31e-263     Length:          4348
Score:                  4609.00       Matches:         901
Percent Similarity:     100.00%       Conservative:    0
Best Local Similarity:  100.00%       Mismatches:      0
Query Match:            100.00%       Indels:          0
DB:                     10            Gaps:            0
```

```
US-10-768-511-8 (1-901) x AAD60813 (1-4348)

Qy          1 MetGluAsnAsnAspAlaLeuAspIleGlyAlaValSerSerProTyrProSerGlnSer 20
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1554 ATGGAAAATAATGATGCACTTGACATTGGAGCCGTGTCGTCCCCATATCCTTCGCAATCT 1613

Qy         21 GluGlyValSerThrProLeuProGlnValThrSerProSerPheAspAsnAlaAlaSer 40
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1614 GAAGGAGTGTCTACGCCATTGCCGCAAGTAACATCACCGAGCTTCGACAATGCAGCCTCA 1673

Qy         41 ProValAlaGlyArgArgAlaValArgGlnThrProThrSerAlaValArgArgArgGly 60
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1674 CCCGTGGCCGGGCGGAGGGCCGTACGGCAGACCCCTACATCTGCAGTTCGAAGGAGAGGG 1733

Qy         61 ArgGluThrAspSerAlaArgArgArgArgSerArgSerArgSerLeuGlyAsnSerVal 80
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1734 AGAGAAACGGATTCCGCTCGTCGTAGGAGGAGTCGATCTCGCAGTTTAGGCAATTCTGTT 1793

Qy         81 TyrSerSerProTyrAspAlaGlyThrProGlyThrProGlyThrProValAlaThrPro 100
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1794 TATAGTTCCCCTTACGATGCGGGGACTCCTGGAACTCCTGGAACTCCAGTGGCTACTCCG 1853

Qy        101 ValTyrAlaThrProValGlyThrProMetGlyThrProSerPheHisArgGlyThrPro 120
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1854 GTTTACGCTACCCCAGTCGGTACACCTATGGGTACCCCATCGTTCCATCGTGGCACGCCA 1913

Qy        121 GlnTyrLysGlnArgSerGluLeuGlySerGlnGlyLysProLeuHisArgArgArgArg 140
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1914 CAGTACAAACAGCGCAGTGAGCTTGGTTCCCAGGGGAAGCCTCTACATCGGAGACGTCGA 1973

Qy        141 SerGlnSerArgGluProGlyHisArgSerProSerArgGluProSerAlaAspGlyArg 160
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1974 TCTCAATCCAGAGAACCCGGGCATCGATCTCCTTCAAGGGAACCTAGTGCTGATGGGCGT 2033

Qy        161 ProSerGluSerAlaGluProAspAspThrLeuGlyGlyGluTyrAlaTyrValTrpGly 180
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       2034 CCCTCTGAATCTGCTGAGCCAGATGACACTTTGGGTGGAGAATATGCTTATGTTTGGGGG 2093

Qy        181 ThrAsnValAsnIleProAspValLeuArgAlaIleArgArgPheLeuHisAsnTyrArg 200
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       2094 ACGAATGTTAACATTCCAGATGTGCTTAGGGCGATTCGTCGATTTCTCCACAATTATCGT 2153

Qy        201 SerSerAlaHisAspLeuAsnSerLysTyrIleGlnIleIleGluGluThrValGluArg 220
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       2154 TCGAGTGCTCATGATCTTAATTCCAAGTACATCCAGATCATAGAGGAGACTGTGGAGCGT 2213

Qy        221 GluGluAspThrLeuAsnIleAspMetSerAspIleTyrAspHisAspProAspLeuTyr 240
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       2214 GAGGAGGATACTCTAAATATCGACATGTCAGACATTTATGACCATGATCCTGATCTATAC 2273

Qy        241 AlaLysIleValArgTyrProLeuAspIleIleProLeuLeuAspThrGluCysGlnGlu 260
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       2274 GCAAAAATTGTTCGATACCCACTCGACATCATCCCCCTGTTGGACACTGAGTGTCAGGAA 2333

Qy        261 ValAlaThrSerLeuLeuProThrPheGluLysHisIleGluAlaArgProPheAsnLeu 280
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       2334 GTTGCTACCTCTTTACTACCAACGTTTGAGAAGCATATTGAGGCCAGACCTTTCAATCTC 2393

Qy        281 LysAlaSerValHisMetArgGluLeuAsnProSerAspIleAspLysLeuValSerVal 300
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       2394 AAAGCATCGGTGCACATGCGTGAACTCAACCCTTCAGATATAGACAAATTGGTTTCTGTT 2453

Qy        301 LysGlyMetValIleArgCysSerSerIleIleProGluIleLysGlyAlaPhePheLys 320
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       2454 AAAGGAATGGTTATCCGGTGCAGTTCTATCATACCTGAAATTAAGGGGGCCTTCTTCAAA 2513

Qy        321 CysLeuValCysGlyHisSerProProLeuValThrValValLysGlyArgValGluGlu 340
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       2514 TGTTTAGTGTGTGGTCACTCGCCTCCGCTAGTTACAGTTGTTAAAGGGCGGGTTGAGGAG 2573

Qy        341 ProThrArgCysGluLysProGluCysAlaAlaArgAsnAlaMetSerLeuIleHisAsn 360
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       2574 CCAACAAGGTGTGAAAAGCCAGAATGTGCAGCACGGAATGCTATGTCTCTTATTCACAAT 2633

Qy        361 ArgCysThrPheAlaAsnLysGlnIleValArgLeuGlnGluThrProAspAlaIlePro 380
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       2634 CGATGCACTTTTGCAAATAAGCAGATAGTGCGTCTTCAAGAAACTCCAGATGCCATTCCT 2693

Qy        381 GluGlyGluThrProHisThrValSerMetCysLeuTyrAsnThrMetValAspAlaVal 400
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       2694 GAAGGAGAGACTCCACACACAGTCAGCATGTGTTTATACAACACTATGGTTGATGCTGTG 2753

Qy        401 LysProGlyAspArgIleGluValThrGlyValPheLysAlaMetAlaValArgValGly 420
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       2754 AAGCCTGGAGATCGTATTGAGGTAACAGGAGTTTTCAAGGCCATGGCAGTTCGAGTTGGT 2813

Qy        421 ProAsnGlnArgThrLeuArgAlaLeuTyrLysThrTyrIleAspCysValHisValLys 440
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       2814 CCGAATCAACGAACATTACGAGCATTGTATAAGACCTACATCGATTGCGTGCACGTCAAG 2873

Qy        441 LysSerAspArgGlyArgLeuGlnThrGluAspProMetGluMetAspLysGluAsnAsp 460
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       2874 AAGTCTGACAGGGGTCGACTGCAAACTGAAGATCCTATGGAGATGGATAAGGAGAATGAT 2933

Qy        461 MetTyrAlaGlyTyrHisGluSerAspThrSerGluAlaAlaAsnGluAlaLysIleGln 480
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       2934 ATGTATGCTGGGTATCATGAAAGTGATACTTCAGAAGCTGCTAATGAAGCAAAGATTCAA 2993

Qy        481 LysLeuLysGluLeuSerLysLeuProGlyIleTyrAspArgLeuSerArgSerLeuAla 500
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       2994 AAACTTAAAGAGCTGTCCAAGCTCCCGGGCATTTATGATAGACTTTCAAGGTCGCTGGCT 3053

Qy        501 ProSerIleTrpGluLeuGluAspIleLysLysGlyLeuLeuCysGlnLeuPheGlyGly 520
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       3054 CCAAGCATTTGGGAGCTTGAAGATATTAAAAAGGGTCTTCTTTGCCAGCTCTTTGGTGGG 3113

Qy        521 LysAlaLysLysIleProSerGlyAlaSerPheArgGlyAspIleAsnValLeuLeuVal 540
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       3114 AAGGCTAAGAAAATTCCATCTGGAGCATCTTTCCGAGGTGACATCAATGTTTTACTTGTT 3173

Qy        541 GlyAspProGlyThrSerLysSerGlnLeuLeuGlnTyrValHisLysIleAlaProArg 560
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       3174 GGGGACCCTGGTACCAGTAAATCTCAGCTGCTTCAGTATGTGCACAAGATAGCTCCTCGT 3233

Qy        561 GlyIleTyrThrSerGlyArgGlySerSerAlaValGlyLeuThrAlaTyrValThrLys 580
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       3234 GGAATCTACACTAGTGGGCGAGGAAGTTCGGCGGTTGGCTGACAGCGTATGTAACGAAG 3293

Qy        581 AspProGluThrArgGluThrValLeuGluSerGlyAlaLeuValLeuSerAspArgGly 600
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       3294 GATCCAGAAACTCGAGAGACGGTATTGGAGAGCGGAGCTTTGGTTCTTAGTGATCGTGGG 3353

Qy        601 IleCysCysIleAspGluPheAspLysMetSerAspAsnAlaArgSerMetLeuHisGlu 620
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       3354 ATATGCTGTATCGATGAGTTCGACAAAATGTCTGATAATGCCCGAAGCATGCTTCATGAG 3413

Qy        621 ValMetGluGlnGlnThrValSerValAlaLysGlyGlyIleIleAlaSerLeuAsnAla 640
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       3414 GTAATGGAGCAACAAACGGTATCTGTAGCCAAAGGGGGTATCATTGCCTCGCTGAACGCT 3473

Qy        641 ArgThrSerValLeuAlaCysAlaAsnProSerGlySerArgTyrAsnAlaArgLeuSer 660
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       3474 CGGACGTCTGTCCTTGCATGTGCAAATCCTAGTGGGTCCCGATACAATGCGCGCCTTTCT 3533

Qy        661 ValIleAspAsnIleGlnLeuProProThrLeuLeuSerArgPheAspLeuIleTyrLeu 680
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       3534 GTGATTGATAACATCCAGCTTCCTCCAACTCTACTTTCTAGATTTGATTTAATTTACTTA 3593

Qy        681 MetLeuAspLysProAspGluGlnAsnAspArgArgLeuAlaArgHisLeuValAlaLeu 700
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       3594 ATGCTCGACAAACCAGACGAGCAAAACGATCGTCGTCTCGCCAGGCATCTCGTGGCTTTA 3653

Qy        701 HisTyrGluAsnTyrGluValSerLysGlnAspAlaLeuAspLeuGlnThrLeuThrAla 720
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       3654 CACTATGAAAACTATGAAGTTTCAAAGCAGGACGCCTTAGATCTACAAAACACTTACCGCG 3713
```

```
Qy        721 TyrIleThrTyrAlaArgGlnHisValHisProThrLeuSerAspGluAlaAlaGluAsp 740
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       3714 TATATCACCTATGCTCGTCAGCATGTACATCCTACATTAAGTGATGAAGCTGCTGAAGAT 3773

Qy        741 LeuIleAsnGlyTyrValGluMetArgGlnLysGlyAsnPheProGlySerSerLysLys 760
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       3774 TTGATTAATGGCTATGTTGAGATGCGCCAAAAGGGCAACTTTCCTGGAAGCAGTAAAAAG 3833

Qy        761 ValIleThrAlaThrProArgGlnLeuGluSerMetIleArgIleSerGluAlaLeuAla 780
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       3834 GTGATAACAGCCACACCTCGGCAACTCGAAAGTATGATTCGTATCAGTGAAGCCCTAGCT 3893

Qy        781 ArgMetArgPheSerGluValValGluLysValAspAlaAlaGluAlaValArgLeuLeu 800
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       3894 CGAATGAGATTTTCTGAAGTGGTAGAGAAAGTTGATGCAGCAGAAGCTGTGCGCCTTTTA 3953

Qy        801 AspValAlaLeuGlnGlnSerAlaThrAspHisAlaThrGlyThrIleAspMetAspLeu 820
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       3954 GACGTCGCTTTGCAGCAATCTGCTACTGATCATGCAACAGGTACGATAGACATGGATCTT 4013

Qy        821 IleThrThrGlyValSerAlaSerGluArgIleArgArgAlaAsnLeuLeuAlaAlaLeu 840
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       4014 ATCACGACTGGAGTGTCGGCCAGCGAGCGTATTCGTCGGGCCAACTTGCTAGCTGCTCTG 4073

Qy        841 ArgGluLeuIleAlaAspLysIleSerProGlySerSerSerGlyLeuLysThrSerGln 860
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       4074 CGAGAGCTTATAGCAGATAAAATTTCACCTGGCAGCTCCTCTGGCTTGAAGACCAGTCAG 4133

Qy        861 LeuLeuGluAspIleArgSerGlnSerSerValAspValSerLeuGlnAspIleLysAsn 880
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       4134 CTTCTTGAGGATATCCGGAGCCAAAGCAGTGTGGACGTTAGTTTGCAGGATATTAAAAAT 4193

Qy        881 AlaLeuGlySerLeuGlnGlyGluGlyPheLeuThrValHisGlyAspIleValLysArg 900
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       4194 GCTCTGGGTAGCCTCCAAGGAGAAGGCTTTCTTACTGTCCATGGTGACATAGTCAAGAGA 4253

Qy        901 Val 901
              |||
Db       4254 GTT 4256
```

RESULT 3
ADX53995

```
ID   ADX53995 standard; cDNA; 2841 BP.
XX
AC   ADX53995;
XX
DT   21-APR-2005  (first entry)
XX
DE   Plant full length insert polynucleotide seqid 28735.
XX
KW   plant protectant; plant growth regulant; gene therapy; plant;
KW   recombinant DNA construct; physical array; plant breeding marker;
KW   cold tolerance; heat tolerance; drought tolerance; herbicide tolerance;
KW   extreme osmotic condition; pathogen tolerance; pest tolerance;
KW   growth rate; cell cycle pathway; disease resistance;
KW   galactomannan production; lignin production; plant growth regulator;
KW   yield; plant growth; plant development; seed oil; protein yield;
KW   protein content; gene; ss.
XX
OS   Unidentified.
XX
PN   US2004034888-A1.
XX
PD   19-FEB-2004.
XX
PF   28-APR-2003; 2003US-00425114.
XX
PR   06-MAY-1999;   99US-00304517.
PR   05-NOV-2001; 2001US-00985678.
XX
PA   (LIUJ/) LIU J.
PA   (ZHOU/) ZHOU Y.
PA   (KOVA/) KOVALIC D K.
PA   (SCRE/) SCREEN S E.
PA   (TABA/) TABASKA J E.
PA   (CAOY/) CAO Y.
XX
PI   Liu J,  Zhou Y,  Kovalic DK,  Screen SE,  Tabaska JE,  Cao Y;
XX
DR   WPI; 2004-180133/17.
XX
PT   New recombinant DNA construct, useful for improving plant tolerance to
PT   cold, heat, drought, herbicides, extreme osmotic conditions, pathogens or
PT   pests, for conferring increased resistance to plant disease, or for
PT   improving yield.
XX
PS   Claim 1; SEQ ID NO 28735; 15pp; English.
XX
CC   The invention describes a recombinant DNA construct comprising a
CC   polynucleotide consisting of a sequence encoding an amino acid sequence
CC   available in electronic form from the US patent office at
CC   ftp.seqdata.uspto.gov/sequence.html?DocID:2004034888. The polynucleotide
CC   of the invention are also useful in physical arrays of molecules and as
CC   plant breeding markers. The recombinant DNA construct is useful for
CC   improving plant tolerance to cold, heat, drought, herbicides, extreme
CC   osmotic conditions, pathogens or pests, for manipulating growth rate in
CC   plant cells by modification of the cell cycle pathway, for conferring
CC   increased resistance to plant disease, for producing galactomannan,
CC   lignin or plant growth regulators, for increasing the rate of homologous
CC   recombination in plants, for improving yield by modification of
CC   photosynthesis or carbohydrate, nitrogen or phosphorus use and/or uptake
CC   or by providing improved plant growth and development under at least one
CC   stress condition or for modifying seed oil or protein yield and/or
CC   content. This sequence represents a plant full length insert
CC   polynucleotide that can be used in the recombinant DNA construct of the
CC   invention.
XX
SQ   Sequence 2841 BP; 707 A; 722 C; 730 G; 682 T; 0 U; 0 Other;
```

```
Alignment Scores:
Pred. No.:               2.77e-141     Length:          2841
Score:                   2550.00       Matches:         514
Percent Similarity:      73.24%        Conservative:    143
Best Local Similarity:   57.30%        Mismatches:      176
Query Match:             55.33%        Indels:          64
DB:                      13            Gaps:            17

US-10-768-511-8 (1-901) x ADX53995 (1-2841)

Qy         20 SerGluGlyValSerThrProLeuProGlnValThrSerProSerPheAspAsnAlaAla 39
              |||:::|||       :::      |||     :::||||||   :::        :::
Db        100 TCTGACGGCGGCGGCAGCTCCCCTCCTCCCGCATCGTCGCCC---TATGGTCGGCCGTCA 156

Qy         40 SerProValAlaGlyArgArgAlaValArgGlnThrProThrSerAlaValArgArgArg 59
              ||||||:::|||                     :::|||:::          ||||||
Db        157 AGCCCACTCGCG---------GTTACTAACTCCTCCCCATCTCAGCCCACACGCCGCCCC 207

Qy         60 GlyArgGluThrAspSerAlaArgArgArgArgSerArgSerArgSerLeuGlyAsnSer 79
              |||               ||||||||||||
Db        208 GGC---------------GGTCGCCGTCGCCGTGGTTCC--------------------- 231

Qy         80 ValTyrSerSerProTyrAspAlaGlyThrProGlyThrProGlyThrProValAlaThr 99
                    :::|||||||||   |||   :::|||       |||
Db        232 ------GCTAGTCCTTAT---GCTTCGTCCCCTTCTCTCGGGGGG--------------- 267

Qy        100 ProValTyrAlaThrProValGlyThrProMetGlyThrProSerPheHisArgGlyThr 119
                    :::   ||||||                     |||       |||   |||
Db        268 ------TTCGAGACGCCG---------------------CCGCACCCTGGCCGCCGCACG 300

Qy        120 ProGlnTyrLysGlnArgSerGluLeuGlySerGlnGlyLysProLeuHisArgArgArg 139
              |||               |||      |||:::   |||   |||      |||:::
Db        301 CCG---------------TCCGGAGCTGGGGCT---GGCGCCCCTCGGCAGCCGCGTCAG 342

Qy        140 ArgSerGlnSerArgGluProGlyHisArgSerProSerArgGluProSerAlaAspGly 159
                 |||      |||   |||         :::||||||         |||   |||
Db        343 AACTCGACTGGCCGGTTCCCG------CCGACGCCCTCCACTCCAATGTCCACCGATGAC 396
```

```
Qy    160 ---ArgProSerGluSerAlaGluProAspAspThrLeuGlyGly--------------- 173
Db    397 GTCCCTCCATCCTCCGAAGCTGGGGACGACGAGACCGACGGCGGCGGCGTCGACGCC 456

Qy    174 GluTyrAlaTyrValTrpGlyThrAsnValAsnIleProAspValLeuArgAlaIleArg 193
Db    457 ACCCCGGTCTTCGTCTGGGGCACCAACATAAGCGTGCAGGACGTCAACGCCGCCATTCTC 516

Qy    194 ArgPheLeuHisAsnTyrArgSerSerAlaHisAspLeu--------------------- 206
Db    517 CGGTTCTTGCGCCACTTCCGG---GACCCGCGCGACGCTGGCCGCGTCGACCCGGTCATG 573

Qy    207 ---AsnSerLysTyrIleGlnIleIleGluGluThrValGluArgGlu---GluAspThr 224
Db    574 GACGAGGGCAAGTACATGCGCGCCATCCACCGCATCCTCGAGCTCGAGGGCGGGGAGTCG 633

Qy    225 LeuAsnIleAspMetSerAspIleTyrAspHisAspProAspLeuTyrAlaLysIleVal 244
Db    634 CTCGACGTCGATGCGCACGACGTGTTTGACCACGACCCAGACCTCTACAGCAAGATGGTT 693

Qy    245 ArgTyrProLeuAspIleIleProLeuLeuAspThrGluCysGlnGluValAlaThrSer 264
Db    694 CGCTATCCGCTCGAGGTGCTCGCCATCTTCGACATCGTGCTCATGGACCTCGTCGCGCGC 753

Qy    265 LeuLeuProThrPheGluLysHisIleGluAlaArgProPheAsnLeuLysAlaSerVal 284
Db    754 ATCGAGCCGCTCTTCGAGAAGCACATCCAGACCAGGATCTACAACCTCAAGTCGTCCATT 813

Qy    285 HisMetArgGluLeuAsnProSerAspIleAspLysLeuValSerValLysGlyMetVal 304
Db    814 TGCTTGAGGAATCTCAACCCATCTGATATTGAGAAGATGGTATCCATCAAGGGTATGATA 873

Qy    305 IleArgCysSerSerIleIleProGluIleLysGlyAlaPhePheLysCysLeuValCys 324
Db    874 ATTAGATGCAGCTCGGTCATACCGGAGCTCAAGGAGGCTGTGTTCCGCTGCCTGGTTTGT 933

Qy    325 GlyHisSerProProLeuValThrValValLysGlyArgValGluGluProThrArgCys 344
Db    934 GGTTTCTACTCAGAGCCCGTCATGGTTGATAGAGGAAGAGTAACTGAACCACACATTTGT 993

Qy    345 GluLysProGluCysAlaAlaArgAsnAlaMetSerLeuIleHisAsnArgCysThrPhe 364
Db    994 CAGAAAGAACAATGTAAAGCCACAAATTCTATGACCCTAGTGCACAACAGATGCAGATTT 1053

Qy    365 AlaAsnLysGlnIleValArgLeuGlnGluThrProAspAlaIleProGluGlyGluThr 384
Db   1054 TCAGACAAGCAGATCATAAAGTTGCAGGAAACACCAGACGAGATACCAGAAGGTGGCACT 1113

Qy    385 ProHisThrValSerMetCysLeuTyrAsnThrMetValAspAlaValLysProGlyAsp 404
Db   1114 CCACATACAGTTAGTGTCTTGATGCATGATAAGCTTGTTGATGCTGGAAAGCCTGGAGAT 1173

Qy    405 ArgIleGluValThrGlyValPheLysAlaMetAlaValArgValGlyProAsnGlnArg 424
Db   1174 AGGGTTGAGATAACTGGAATATACAGAGCTATGAGTATTCGAATTGGACCAACTCAAAGG 1233

Qy    425 ThrLeuArgAlaLeuTyrLysThrTyrIleAspCysValHisValLysLysSerAspArg 444
Db   1234 ACAGTGAAGTCTATATTCAAGACATATATTGATTGCCTTCACATAAAGAAGACAGACAAG 1293

Qy    445 GlyArgLeuGlnThrGluAspProMetGluMetAspLysGluAsnAspMetTyrAlaGly 464
Db   1294 TCTAGGCTTCATGTGGAGGACACCATGGATATTGATAATTCTAAC--------------- 1338

Qy    465 TyrHisGluSerAspThrSerGluAlaAlaAsnGluAlaLysIleGlnLysLeuLysGlu 484
Db   1339 ---GCTAGCAAATCTACTGAAGAGGATTTTCTTAGTGATAAGGTTGAGAAACTAAAAGAG 1395

Qy    485 LeuSerLysLeuProGlyIleTyrAspArgLeuSerArgSerLeuAlaProSerIleTrp 504
Db   1396 CTTTCGAAGTTGCCTGATATCTATGAAAGATTGACTAGATCATTAGCTCCAAACATATGG 1455

Qy    505 GluLeuGluAspIleLysLysGlyLeuLeuCysGlnLeuPheGlyGlyLysAlaLysLys 524
Db   1456 GAGTTGGATGATGTCAAAAGAGGTCTCCTTTGCCAGCTTTTCGGCGGTAATCCCTTGAAG 1515

Qy    525 IleProSerGlyAlaSerPheArgGlyAspIleAsnValLeuLeuValGlyAspProGly 544
Db   1516 CTTCCTTCTGGAGCTAGTTTCCGGGGTGACATCAATATTTTACTTGTGGGGGACCCTGGA 1575

Qy    545 ThrSerLysSerGlnLeuLeuGlnTyrValHisLysIleAlaProArgGlyIleTyrThr 564
Db   1576 ACAAGTAAATCCCAGCTTCTCCAGTACATGCATAAACTGTCTCCTCGTGGTATCTATACG 1635

Qy    565 SerGlyArgGlySerSerAlaValGlyLeuThrAlaTyrValThrLysAspProGluThr 584
Db   1636 AGTGGTAGAGGAAGTTCTGCTGTTGGTCTTACTGCTTATGTTACCAAAGACCCTGAGACT 1695

Qy    585 ArgGluThrValLeuGluSerGlyAlaLeuValLeuSerAspArgGlyIleCysCysIle 604
Db   1696 GGCGAAACTGTTCTAGAAAGTGGAGCACTTGTTTTGAGTGACAAAGGTGTTTGTTGCATA 1755

Qy    605 AspGluPheAspLysMetSerAspAsnAlaArgSerMetLeuHisGluValMetGluGln 624
Db   1756 GATGAGTTTGATAAGATGTCTGATAATGCCCGAAGCATGTTACACGAGGTGATGGAACAG 1815

Qy    625 GlnThrValSerValAlaLysGlyGlyIleIleAlaSerLeuAsnAlaArgThrSerVal 644
Db   1816 CAGACAGTATCCATTGCGAAGGCTGGAATAATTGCATCTTTAAACGCTAGGACATCTGTC 1875

Qy    645 LeuAlaCysAlaAsnProSerGlySerArgTyrAsnAlaArgLeuSerValIleAspAsn 664
Db   1876 CTGGCATGTGCCAATCCTACTGAATCACGTTACAATCCAAGGCTCTCTGTAATTGACAAC 1935

Qy    665 IleGlnLeuProProThrLeuLeuSerArgPheAspLeuIleTyrLeuMetLeuAspLys 684
Db   1936 ATCCACTTAGCGCCAACGCTACTTTCAAGATTCGACCTGATTTATCTTATCTTGGACAAG 1995

Qy    685 ProAspGluGlnAsnAspArgArgLeuAlaArgHisLeuValAlaLeuHisTyrGluAsn 704
Db   1996 GCGGATGAGCAAACTGATAGGCGCCTGGCAAAGCATATTGTTTCGTTGCATTTTGAGAAT 2055

Qy    705 TyrGluValSerLysGlnAspAlaLeuAspLeuGlnThrLeuThrAlaTyrIleThrTyr 724
Db   2056 CCAAATTTAGAGGAGCTCGAGGTCTTGGACTTGCAGACACTAGTTTCCTACATAAGCTAT 2115

Qy    725 AlaArgGlnHisValHisProThrLeuSerAspGluAlaAlaGluAspLeuIleAsnGly 744
Db   2116 GCAAGGAAGTATATTCAGCCACAGTTATCTGATGAAGCTGCAGAAGAGTTAACTCGTGGC 2175

Qy    745 TyrValGluMetArgGlnLysGlyAsnPheProGlySerSerLysLysValIleThrAla 764
Db   2176 TATGTCGAGATGAGAAAAAGAGGGAATAGCCCTGGGAGCAGAAAGAAGGTCATAACAGCA 2235

Qy    765 ThrProArgGlnLeuGluSerMetIleArgIleSerGluAlaLeuAlaArgMetArgPhe 784
Db   2236 ACCGCTAGACAAATAGAGAGTTTGATCCGTCTCAGCGAAGCATTAGCCCGAATGCGGTTC 2295

Qy    785 SerGluValValGluLysValAspAlaAlaGluAlaValArgLeuLeuAspValAlaLeu 804
Db   2296 TCTGAAGTGGTCGAGGTGCGGGATGTTGTGGAGGCATTCAGGCTTCTTGAAGTCGCCATG 2355

Qy    805 GlnGlnSerAlaThrAspHisAlaThrGlyThrIleAspMetAspLeuIleThrThrGly 824
Db   2356 CAGCAGTCTGCGACGGATCATGCAACTGGTACGATTGATATGGATCTGATCATGACGGGG 2415

Qy    825 ValSerAlaSerGluArgIleArgArgAlaAsnLeuLeuAlaAlaLeuArgGluLeuIle 844
Db   2416 ATATCCGCAAGCGAAAGGCAGAGGCGGGAGAACCTCGTTGCCGCAACCCGTAACCTGATT 2475

Qy    845 AlaAspLysIleSerProGlySerSerSerGlyLeuLysThrSerGlnLeuLeuGluAsp 864
Db   2476 GCGGAGAAAATGCAGCTTGGAGGCCCCTCG---ATGCGCATGATTGAGTTGCTGGAGGAA 2532

Qy    865 IleArgSerGlnSerSerValAspValSerLeuGlnAspIleLysAsnAlaLeuGlySer 884
```

```
                :::|||  ||||||||::::::::   :::   :::::::::  |||||||||:::
Db       2533 CTGAGGAAGCAGAGCTCAATGGAAATTCATATGCACGAACTCCGCGGTGCTCTTGGCACC 2592

Qy        885 LeuGlnGlyGluGlyPheLeuThrValHisGlyAspIleValLysArgVal 901
              |||      |||||   :::   :::|||||||||  |||:::||||||
Db       2593 CTGATGACTGAAGGCGCGGTGGTTATCCATGGAGACAACGTGAGGAGAGTT 2643

RESULT 4
ABT21208
ID   ABT21208 standard; DNA; 3072 BP.
XX
AC   ABT21208;
XX
DT   16-APR-2003  (first entry)
XX
DE   Aspergillus fumigatus essential gene #3566.
XX
KW   Fungicide; cytostatic; essential gene; Aspergillus fumigatus; infection;
KW   cancer; contamination; biofilm; antibody; immune response; ds.
XX
OS   Aspergillus fumigatus.
XX
PN   WO200286090-A2.
XX
PD   31-OCT-2002.
XX
PF   23-APR-2002; 2002WO-US013142.
XX
PR   23-APR-2001; 2001US-0285697P.
PR   27-APR-2001; 2001US-0287066P.
PR   05-JUN-2001; 2001US-0295890P.
PR   09-JUL-2001; 2001US-0303899P.
PR   31-AUG-2001; 2001US-0316362P.
XX
PA   (ELIT-) ELITRA PHARM INC.
XX
PI   Jiang B,  Tishkoff D,  Zamudio C,  Eroshkin AM,  Hu W,  Lemieux SM;
XX
DR   WPI; 2003-093124/08.
XX
PT   New purified or isolated nucleic acids of essential genes of Aspergillus
PT   fumigatus, useful for treating or preventing infections by A. fumigatus,
PT   or for treating a non-infectious disease in a subject e.g. cancer.
XX
PS   Disclosure; Page; 175pp; English.
XX
CC   The invention relates to novel purified or isolated nucleic acids of
CC   essential genes of Aspergillus fumigatus. The isolated nucleic acids of
CC   the invention are used to treat or prevent infections by a pathogenic
CC   organism such as A. fumigatus, to treat a non-infectious disease in a
CC   subject (e.g. cancer), to prevent or contain contamination of an object
CC   by A. fumigatus, or to prevent or inhibit formation on a surface of a
CC   biofilm comprising A. fumigatus. The polynucleotides are useful for
CC   expressing recombinant protein for characterisation, screening or
CC   therapeutic use, as markers for host tissues in which the pathogenic
CC   organisms invade or reside, for comparing with the DNA sequence of A.
CC   fumigatus to identify duplicated genes or paralogues having the same or
CC   similar biochemical activity and/or function, for comparing with DNA
CC   sequences of other related or distant pathogenic organisms to identify
CC   potential orthologous essential or virulence genes, for selecting and
CC   making oligomers for attachment to a nucleic acid array for examination
CC   of expression patterns, for raising anti-protein antibodies, as an
CC   antigen to raise anti-DNA antibodies or to elicit another immune
CC   response, and for identifying polynucleotides encoding the other protein
CC   with which binding occurs or to identify inhibitors of the binding
CC   interaction. The polypeptides may be used to raise antibodies or to
CC   elicit immune response, as a reagent in assays designed to quantitatively
CC   determine levels of the protein in biological fluids, as a marker for
CC   host tissues in which pathogenic organism invade or reside, and to
CC   isolate correlative receptors or ligands in the case or virulence
CC   factors. This polynucleotide sequence represents one of the essential
CC   genes of Aspergillus fumigatus of the invention
XX
SQ   Sequence 3072 BP; 717 A; 847 C; 827 G; 681 T; 0 U; 0 Other;

Alignment Scores:
Pred. No.:              8e-102         Length:         3072
Score:                  1883.00        Matches:        438
Percent Similarity:     57.77%         Conservative:   157
Best Local Similarity:  42.52%         Mismatches:     278
Query Match:            40.85%         Indels:         159
DB:                     8              Gaps:           23

US-10-768-511-8 (1-901) x ABT21208 (1-3072)

Qy         12 ValSerSerProTyrProSerGlnSerGluGlyValSerThrProLeuProGlnValThr 31
              :::|||||||||        |||:::        |||
Db          1 ATGTCTTCTCCAGCTTCTTCTCGCCGTAGGGGC--------------------------- 33

Qy         32 SerProSerPheAspAsnAlaAlaSerProValAlaGlyArgArgAlaValArgGlnThr 51
                 |||:::   |||  |||   |||      |||             ||||||
Db         34 CGCCCAGCCAAGGACACGGCAACCTCTTCTCCCGCGCGGTCAACCCGTTCTCGTCAGCTC 93

Qy         52 ProThrSerAlaValArgArgArgGlyArgGluThrAspSerAlaArgArgArgArgSer 71
                 ||||||  ||||||   :::||||||:::   :::|||   |||||||||
Db         94 CAGACCAGTA--GTCCGACTCCAAGGGCGGCAGATGAACAGTCACAGGCGACGCCT-CGT 150

Qy         72 ArgSerArgSerLeuGlyAsnSerVal-----------------------------Tyr 81
                 ||||||   |||
Db        151 GCTTCAAGGCGACTGAGAGGCGAAGCGGCAGTTCCCTCGTCATCGCCTATGTTCTTCCAG 210

Qy         82 SerSerProTyrAspAlaGlyThrProGly---ThrProGlyThrProValAlaThrPro 100
              |||||||||      |||   :::           ||||||         :::   |||
Db        211 TCCTCCCCTTCAAAGGCAGACAGCAGTGCGGAAACACCAGATGTTCGAATGGATGAGCCC 270

Qy        101 ValTyrAlaThrProVal-----------------------------GlyThrProMet 110
                    ::::::|||:::                                 ||||||
Db        271 ------AGTTCACCAATGCGAGAATCTTCTACTATGGATGAAGGAGACAGGACACCTAGA 324

Qy        111 Gly---ThrProSerPheHisArgGlyThrPro-------------------------- 120
              |||      |||:::              :::|||
Db        325 GGAAATGCTCCTACCATGAGAGACTCCTCCCCCATCCGGTATATGTCTAGCTCCAGTCCA 384

Qy        121 ---------------GlnTyrLysGlnArgSerGluLeuGlySerGlnGlyLysProLeu 135
                             :::      :::|||||||::::::    |||          |||
Db        385 ACTCGAGCTCAGAATCGCCAGTCCCGGCGTTCCGACATTCCCAGCAGCAGCAGCGGATTA 444

Qy        136 His------------------------------------ArgArgArgArgSerGlnSer 143
                                                        |||||||||
Db        445 TTCGTCTCGTCCAGACCCAGCATCGAGAGCAACCGTGCCGTGTCCCGCCGTAGCGACCTT 504

Qy        144 ArgGluProGlyHis---ArgSerProSerArgGlu----------------------- 154
                       |||         ||||||:::|||
Db        505 CATTCTGGTGGTTTTCTTTCTAGCCCGAACCGTCGCCGCAGAGTTTTTGTCGATGCTAAT 564

Qy        155 ------ProSerAlaAspGlyArgProSerGluSerAla--------------GluPro 167
                    |||:::   ||||||  |||        |||                 |||
Db        565 GGTATGCCTGCGACCGATGGCGATCCACGTTCCGATGCCACCTTCTCGAATATCCACCCA 624

Qy        168 Asp------------AspThrLeuGlyGlyGluTyrAla---TyrValTrpGlyThrAsn 182
              |||            :::   |||||||||                :::||||||||||||
Db        625 GACACCTCTGAGGCCGAGGCCTTGGGCGGTAGCTCAACCCGTGTGATTTGGGGTACCAAC 684

Qy        183 ValAsnIleProAspValLeuArgAlaIleArgArgPheLeuHisAsn------------ 198
              ::::::|||   |||   :::   |||   :::   ||||||:::|||
Db        685 ATCTCCATCCAGGATTCCATGTCCGCATTCAAGAACTTTCTCTACAACTTCCAAACAAAA 744

Qy        199 TyrArgSerSerAlaHisAspLeuAsnSerLysTyrIleGlnIleIleGluGluThrVal 218
              ||||||        |||                       :::|||:::   ::::::
Db        745 TACCGTCTGTGGGCAGAGGGCGCAACCGAGGATGAGACACGTATAATGGGTGATTCAGCG 804

Qy        219 GluArgGluGlu-----------------------------------------AspThrLeu 225
              |||      |||                                             :::|||
Db        805 GAGGAACGGGAATACATCAGCATGTTGAGCACCATGCGGCAACTTGGAGTAACCAGCTTG 864

Qy        226 AsnIleAspMetSerAspIle------TyrAspHisAspProAspLeuTyrAlaLysIle 243
```

```
Db    865 AACTTGGATGCGAAGAACCTGAAGGCATACCCATCAACACTCAAGCTATGGCATCAGTTG 924

Qy    244 ValArgTyrProLeuAspIleIleProLeuLeuAspThrGluCysGlnGluValAlaThr 263

Db    925 CATGCTTATCCTCAAGAGATCATTCCATTGATGGATCAGACAGTGAAGGATGTGATGGTG 984

Qy    264 SerLeu------------------------------------------------------ 265

Db    985 GAGCTTGCGATCAAGGAGATGGAGCGCCTGCGGGCTCAGAACCAACGAAACCAGAATCAT 1044

Qy    265 ------------------------------------------------------------ 265

Db   1045 AACAGAGGCCTGAGCTCTGGTCCCGCTGTTCCCAGCTCTGACGCGCTGAGTGAAACTGGC 1104

Qy    266 ------------------LeuProThrPheGluLysHisIleGluAlaArg--------- 276

Db   1105 AGAATGCCACAGAACGAGATCCCCGACCTTGTCGGTGAAGTGGAAACCAAGGCATTCAAA 1164

Qy    277 ------ProPheAsnLeuLysAlaSerValHisMetArgGluLeuAsnProSerAspIle 294

Db   1165 GTTCTGCCTTTTGGTCTCGACTCGACTGTGAATATGAGAGACCTCGATCCTGCAGACATG 1224

Qy    295 AspLysLeuValSerValLysGlyMetValIleArgCysSerSerIleIleProGluIle 314

Db   1225 GATAAACTAGTAAGCATTAAGGGCTTAGTCATTCGAACGACACCCATCATTCCTGATATG 1284

Qy    315 LysGlyAlaPhePheLysCysLeuValCysGlyHisSerProProLeuValThrValVal 334

Db   1285 AAAGAAGCTTTCTTCCGTTGCCAAGTCTGCAACCATGGTGTTCAG---GTCGACATTGAT 1341

Qy    335 LysGlyArgValGluGluProThrArgCysGluLysProGluCysAlaAlaArgAsnAla 354

Db   1342 CGTGGAAAGATTGCGGAACCCACTGAGTGCCCACGTCCAGTGTGTAAGGAACGAAACTCG 1401

Qy    355 MetSerLeuIleHisAsnArgCysThrPheAlaAsnLysGlnIleValArgLeuGlnGlu 374

Db   1402 ATGCAACTCATCCATAACCGCTGTGTATTTGCCGACAAGCAGGTCATCAAGTTGCAGGAA 1461

Qy    375 ThrProAspAlaIleProGluGlyGluThrProHisThrValSerMetCysLeuTyrAsn 394

Db   1462 ACACCTGACAGCATTCCTGATGGCCAGACTCCTCACTCGGTTTCCCTTTGTGTGTATGAT 1521

Qy    395 ThrMetValAspAlaValLysProGlyAspArgIleGluValThrGlyValPheLysAla 414

Db   1522 GAGCTGGTGGATGTCTGCAAGGCTGGTGATCGGGTCGAAGTGACCGGTATTTTCCGGTGC 1581

Qy    415 MetAlaValArgValGlyProAsnGlnArgThrLeuArgAlaLeuTyrLysThrTyrIle 434

Db   1582 AACCCTGTGCGCGTTAATCCTCGCCAGCGTACACAGAAGTCGCTGTTCAAGACGTACATA 1641

Qy    435 AspCysValHisValLysLysSerAspArgGlyArgLeuGlnThrGluAspProMetGlu 454

Db   1642 GATGTTCTTCATGTTCAGAAGATCGATCGCAAGAAGTTGGGTATCGACGTCTCGACCATC 1701

Qy    455 MetAspLysGluAsnAspMetTyrAlaGlyTyrHisGluSerAspThrSerGluAlaAla 474

Db   1702 GAGCAGGAGCTCTCGGAACAGGCGGCTGGGGATGCAGAACAAACACGTAGGCTCACTGCG 1761

Qy    475 AsnGluAlaLysIleGlnLysLeuLysGluLeuSerLysLeuProGlyIleTyrAspArg 494

Db   1762 GAGGAGGAA------GAGAAAATTAAGCGAACTGCTACCAGACCTGATCTGTATGAGCTT 1815

Qy    495 LeuSerArgSerLeuAlaProSerIleTrpGluLeuGluAspIleLysLysGlyLeuLeu 514

Db   1816 CTCTCTCGGTCCTTGGCCCCCAGCATCTACGAGATGGACGACGTGAAGAAGGGAATCCTG 1875

Qy    515 CysGlnLeuPheGlyGlyLysAlaLysLysIleProSerGlyAlaSer------PheArg 532

Db   1876 CTTCAGTTGTTTGGAGGCACCAACAAGACCTTCCAGAAGGGTGGTAACCCACGATACCGT 1935

Qy    533 GlyAspIleAsnValLeuLeuValGlyAspProGlyThrSerLysSerGlnLeuLeuGln 552

Db   1936 GGAGATATCAATATCCTTCTCTGTGGTGACCCATCTACATCCAAGTCCCAGCTTCTTCGT 1995

Qy    553 TyrValHisLysIleAlaProArgGlyIleTyrThrSerGlyArgGlySerSerAlaVal 572

Db   1996 TACGTCCATAAGATTGCCCCTCGCGGTGTGTATACCAGCGGCAAGGGCTCCTCGGCTGTT 2055

Qy    573 GlyLeuThrAlaTyrValThrLysAspProGluThrArgGluThrValLeuGluSerGly 592

Db   2056 GGTCTTACGGCGTACGTCACCCGCGATCCTGAAACCCGCCAGATGGTCCTCGAGTCGGGT 2115

Qy    593 AlaLeuValLeuSerAspArgGlyIleCysCysIleAspGluPheAspLysMetSerAsp 612

Db   2116 GCCTTGGTTCTTTCAGACGGCGGTATCTGTTGCATCGACGAGTTCGACAAGATGAACGAA 2175

Qy    613 AsnAlaArgSerMetLeuHisGluValMetGluGlnGlnThrValSerValAlaLysGly 632

Db   2176 TCCACTCGGTCCGTTCTGCATGAAGTCATGGAACAACAGACAGTATCTATCGCCAAGGCA 2235

Qy    633 GlyIleIleAlaSerLeuAsnAlaArgThrSerValLeuAlaCysAlaAsnProSerGly 652

Db   2236 GGCATTATCACTACTTTGAACGCTAGGACCAGCATCCTGGCTTCCGCCAATCCGATCGGT 2295

Qy    653 SerArgTyrAsnAlaArgLeuSerValIleAspAsnIleGlnLeuProProThrLeuLeu 672

Db   2296 AGCAGGTACAATCCCAACTTGCCCGTTCCTCAAAATATTGACCTTCCGCCTACCTTGCTC 2355

Qy    673 SerArgPheAspLeuIleTyrLeuMetLeuAspLysProAspGluGlnAsnAspArgArg 692

Db   2356 TCCCGATTCGACTTGGTATACCTCGTGCTGGACCGAGTGGATGAGCAGGAAGATCGTCGG 2415

Qy    693 LeuAlaArgHisLeuValAlaLeuHisTyrGluAsn------TyrGluValSerLysGln 710

Db   2416 CTCGCTAAGCATCTTGTCAATATGTACCTGGAAGACAGACCTGAGCATGCTGCCGAGCAA 2475

Qy    711 AspAlaLeuAspLeuGlnThrLeuThrAlaTyrIleThrTyrAlaArgGlnHisValHis 730

Db   2476 GAAATCTTGCCGATCGAATTCCTTACAGCCTATATCACCTACGCCAAGACCAAAGTCCAT 2535

Qy    731 ProThrLeuSerAspGluAlaAlaGluAspLeuIleAsnGlyTyrValGluMetArgGln 750

Db   2536 CCAGTGCTCACACCGGCCGCCGGTAAAGCCTTGTCGGATGCTTACGTTAACATGCGTAAG 2595

Qy    751 LysGlyAsnPheProGlySerSerLysLysValIleThrAlaThrProArgGlnLeuGlu 770

Db   2596 CTTGGAGATGACATCCGGTCTTCTGACCGCCGTATCACCGCTACCACTCGTCAACTGGAG 2655

Qy    771 SerMetIleArgIleSerGluAlaLeuAlaArgMetArgPheSerGluValValGluLys 790

Db   2656 TCCATGATCCGACTGTCGGAAGCGCATGCGCGTATGCGGCTATCGCCGGAGGTCACTGCG 2715

Qy    791 ValAspAlaAlaGluAlaValArgLeuLeuAspValAlaLeuGlnGlnSerAlaThrAsp 810

Db   2716 GATGATGTGGAGGAAGCCGTGCGCCTGATCCGCTCCGCCATCAAGCAGGCGGCCACTGAC 2775

Qy    811 HisAlaThrGlyThrIleAspMetAspLeuIleThrThrGlyValSerAlaSerGluArg 830

Db   2776 TCTCGGACCGGTCTGATCGACATGAGCTTGTTGACGGAGGGCACTAGTGCCAGCGAGAGA 2835

Qy    831 IleArgArgAlaAsnLeuLeuAlaAlaLeuArgGluLeuIleAlaAspLysIleSerPro 850

Db   2836 CGCAGCCGGGAAGCACTCAAGCGTGCCTTGCTGTCTGTGGTGGATGATCTGTGCAGCGGT 2895

Qy    851 GlySerSerSerGlyLeuLysThrSerGlnLeuLeuGluAspIleArgSerGlnSerSer 870

Db   2896 GGCGGTGCAGCT------CGCTGGGCCGAGGTCTTCAGGATCTTAAGCGAGAACAGCAGC 2949

Qy    871 ValAspValSerLeuGlnAspIleLysAsnAlaLeuGlySerLeuGlnGlyGluGlyPhe 890

Db   2950 ATTGAGGTGGATGGAGCCCAGTTTGCGGATGCGGTTCGAGCGCTGGAGGCTGAGGGAGCG 3009

Qy    891 LeuThrValHisGlyAspIleValLysArg 900

Db   3010 GTGAGTGTGGTCGGCGAGGGTGCGCGGCGG 3039
```

```
RESULT 5
ABT19388
ID   ABT19388 standard; DNA; 2625 BP.
XX
AC   ABT19388;
XX
DT   16-APR-2003  (first entry)
XX
DE   Aspergillus fumigatus essential gene #1746.
XX
KW   Fungicide; cytostatic; essential gene; Aspergillus fumigatus; infection;
KW   cancer; contamination; biofilm; antibody; immune response; ds.
XX
OS   Aspergillus fumigatus.
XX
PN   WO200286090-A2.
XX
PD   31-OCT-2002.
XX
PF   23-APR-2002; 2002WO-US013142.
XX
PR   23-APR-2001; 2001US-0285697P.
PR   27-APR-2001; 2001US-0287066P.
PR   05-JUN-2001; 2001US-0295890P.
PR   09-JUL-2001; 2001US-0303899P.
PR   31-AUG-2001; 2001US-0316362P.
XX
PA   (ELIT-) ELITRA PHARM INC.
XX
PI   Jiang B,  Tishkoff D,  Zamudio C,  Eroshkin AM,  Hu W,  Lemieux SM;
XX
DR   WPI; 2003-093124/08.
XX
PT   New purified or isolated nucleic acids of essential genes of Aspergillus
PT   fumigatus, useful for treating or preventing infections by A. fumigatus,
PT   or for treating a non-infectious disease in a subject e.g. cancer.
XX
PS   Disclosure; Page; 175pp; English.
XX
CC   The invention relates to novel purified or isolated nucleic acids of
CC   essential genes of Aspergillus fumigatus. The isolated nucleic acids of
CC   the invention are used to treat or prevent infections by a pathogenic
CC   organism such as A. fumigatus, to treat a non-infectious disease in a
CC   subject (e.g. cancer), to prevent or contain contamination of an object
CC   by A. fumigatus, or to prevent or inhibit formation on a surface of a
CC   biofilm comprising A. fumigatus. The polynucleotides are useful for
CC   expressing recombinant protein for characterisation, screening or
CC   therapeutic use, as markers for host tissues in which the pathogenic
CC   organisms invade or reside, for comparing with the DNA sequence of A.
CC   fumigatus to identify duplicated genes or paralogues having the same or
CC   similar biochemical activity and/or function, for comparing with DNA
CC   sequences of other related or distant pathogenic organisms to identify
CC   potential orthologous essential or virulence genes, for selecting and
CC   making oligomers for attachment to a nucleic acid array for examination
CC   of expression patterns, for raising anti-protein antibodies, as an
CC   antigen to raise anti-DNA antibodies or to elicit another immune
CC   response, and for identifying polynucleotides encoding the other protein
CC   with which binding occurs or to identify inhibitors of the binding
CC   interaction. The polypeptides may be used to raise antibodies or to
CC   elicit immune response, as a reagent in assays designed to quantitatively
CC   determine levels of the protein in biological fluids, as a marker for
CC   host tissues in which pathogenic organism invade or reside, and to
CC   isolate correlative receptors or ligands in the case or virulence
CC   factors. This polynucleotide sequence represents one of the essential
CC   genes of Aspergillus fumigatus of the invention
XX
SQ   Sequence 2625 BP; 614 A; 704 C; 713 G; 594 T; 0 U; 0 Other;

Alignment Scores:
Pred. No.:            3.18e-101     Length:         2625
Score:                1871.50       Matches:        410
Percent Similarity:   63.72%        Conservative:   145

Best Local Similarity:  47.07%      Mismatches:   239
Query Match:            40.61%      Indels:       77
DB:                     8           Gaps:         16

US-10-768-511-8 (1-901) x ABT19388 (1-2625)

Qy     89 ThrProGlyThrProGlyThrProValAlaThrProValTyrAlaThrProValGlyThr 108
Db     38 AGTCCCGGCGTTCCGACATTCCCAGCAGCAGCAGCGGATTATTCGTCT------------ 85

Qy    109 ProMetGlyThrPro-SerPheHisArgGlyThrPro-GlnTyrLysGlnArgSerGluL 128
Db     86 ---------CGTCCAGACCCAGCATCGAGAGCAACCGTGCCGTGTCCCGCCGTAGCGACC 136

Qy    128 euGlySerGlnGly------LysProLeuHisArgArgArgArgSerGlnSerArgGluP 146
Db    137 TTCATTCTGGTGGTTTTCTTTCTAGCCCGAACCGTCGCCGCAGAGTTTTTGTCGATGCTA 196

Qy    146 roGlyHisArgSerProSerArgGluProSerAlaAspGlyArgProSerGluSerAla- 165
Db    197 ATGGT-----------------ATGCCTGCGACCGATGGCGATCCACGTTCCGATGCCA 238

Qy    166 -------------GluProAsp-----------AspThrLeuGlyGlyGluTyrAla- 176
Db    239 CCTTCTCGAATATCCACCCAGACACCTCTGAGGCCGAGGCCTTGGGCGGTAGCTCAACCC 298

Qy    177 --TyrValTrpGlyThrAsnValAsnIleProAspValLeuArgAlaIleArgArgPheL 196
Db    299 GTGTGATTTGGGGTACCAACATCTCCATCCAGGATTCCATGTCCGCATTCAAGAACTTTC 358

Qy    196 euHisAsn-----------TyrArgSerSerAlaHisAspLeuAsnSerLysTyrIleG 212
Db    359 TCTACAACTTCCAAACAAAATACCGTCTGTGGGCAGAGGGCGCAACCGAGGATGAGACAC 418

Qy    212 lnIleIleGluGluThrValGluArgGluGlu-------------------------- 222
Db    419 GTATAATGGGTGATTCAGCGGAGGAACGGGAATACATCAGCATGTTGAGCACCATGCGGC 478

Qy    223 -----------AspThrLeuAsnIleAspMetSerAspIle------TyrAspHisAspP 237
Db    479 AACTTGGAGTAACCAGCTTGAACTTGGATGCGAAGAACCTGAAGGCATACCCATCAACAC 538

Qy    237 roAspLeuTyrAlaLysIleValArgTyrProLeuAspIleIleProLeuLeuAspThrG 257
Db    539 TCAAGCTATGGCATCAGTTGCATGCTTATCCTCAAGAGATCATTCCATTGATGGATCAGA 598

Qy    257 luCysGlnGluValAlaThrSerLeu-------------------------------- 265
Db    599 CAGTGAAGGATGTGATGGTGGAGCTTGCGATCAAGGAGATGGAGCGCCTGCGGGCTCAGA 658

Qy    266 ----------------------------LeuProThrPheGlu---LysHisIleGluA 275
Db    659 ACCAACGAAACCAGAATCATAACAGAGGCCTGAGCTCTGTGGAAACCAAGGCATTCAAAG 718

Qy    275 laArgProPheAsnLeuLysAlaSerValHisMetArgGluLeuAsnProSerAspIleA 295
Db    719 TTCTGCCTTTTGGTCTCGACTCGACTGTGAATATGAGAGACCTCGATCCTGCAGACATGG 778

Qy    295 spLysLeuValSerValLysGlyMetValIleArgCysSerSerIleIleProGluIleL 315
Db    779 ATAAACTAGTAAGCATTAAGGGCTTAGTCATTCGAACGACACCCATCATTCCTGATATGA 838

Qy    315 ysGlyAlaPhePheLysCysLeuValCysGlyHisSerProProLeuValThrValValL 335
Db    839 AAGAAGCTTTCTTCCGTTGCCAAGTCTGCAACCATGGTGTTCAG---GTCGACATTGATC 895

Qy    335 ysGlyArgValGluGluProThrArgCysGluLysProGluCysAlaAlaArgAsnAlaM 355
Db    896 GTGGAAAGATTGCGGAACCCACTGAGTGCCCACGTCCAGTGTGTAAGGAACGAAACTCGA 955

Qy    355 etSerLeuIleHisAsnArgCysThrPheAlaAsnLysGlnIleValArgLeuGlnGluT 375
Db    956 TGCAACTCATCCATAACCGCTGTGTATTTGCCGACAAGCAGGTCATCAAGTTGCAGGAAA 1015
```

```
Qy        375 hrProAspAlaIleProGluGlyGluThrProHisThrValSerMetCysLeuTyrAsnT 395
Db       1016 CACCTGACAGCATTCCTGATGGCCAGACTCCTCACTCGGTTTCCCTTTGTGTGTATGATG 1075

Qy        395 hrMetValAspAlaValLysProGlyAspArgIleGluValThrGlyValPheLysAlaM 415
Db       1076 AGCTGGTGGATGTCTGCAAGGCTGGTGATCGGGTCGAAGTGACCGGTATTTTCCGGTGCA 1135

Qy        415 etAlaValArgValGlyProAsnGlnArgThrLeuArgAlaLeuTyrLysThrTyrIleA 435
Db       1136 ACCCTGTGCGCGTTAATCCTCGCCAGCGTACACAGAAGTCGCTGTTCAAGACGTACATAG 1195

Qy        435 spCysValHisValLysLysSerAspArgGlyArgLeuGlnThrGluAspProMetGluM 455
Db       1196 ATGTTCTTCATGTTCAGAAGATCGATCGCAAGAAGTTGGGTATCGACGTCTCGACCATCG 1255

Qy        455 etAspLysGluAsnAspMetTyrAlaGlyTyrHisGluSerAspThrSerGluAlaAlaA 475
Db       1256 AGCAGGAGCTCTCGGAACAGGCGGCTGGGGATGCAGAACAAACACGTAGGCTCACTGCGG 1315

Qy        475 snGluAlaLysIleGlnLysLeuLysGluLeuSerLysLeuProGlyIleTyrAspArgL 495
Db       1316 AGGAGGAA------GAGAAAATTAAGCGAACTGCTACCAGACCTGATCTGTATGAGCTTC 1369

Qy        495 euSerArgSerLeuAlaProSerIleTrpGluLeuGluAspIleLysLysGlyLeuLeuC 515
Db       1370 TCTCTCGGTCCTTGGCCCCCAGCATCTACGAGATGGACGACGTGAAGAAGGGAATCCTGC 1429

Qy        515 ysGlnLeuPheGlyGlyLysAlaLysLysIleProSerGlyAlaSer------PheArgG 533
Db       1430 TTCAGTTGTTTGGAGGCACCAACAAGACCTTCCAGAAGGGTGGTAACCCACGATACCGTG 1489

Qy        533 lyAspIleAsnValLeuLeuValGlyAspProGlyThrSerLysSerGlnLeuLeuGlnT 553
Db       1490 GAGATATCAATATCCTTCTCTGTGGTGACCCATCTACATCCAAGTCCCAGCTTCTTCGTT 1549

Qy        553 yrValHisLysIleAlaProArgGlyIleTyrThrSerGlyArgGlySerSerAlaValG 573
Db       1550 ACGTCCATAAGATTGCCCCTCGCGGTGTGTATACCAGCGGCAAGGGCTCCTCGGCTGTTG 1609

Qy        573 lyLeuThrAlaTyrValThrLysAspProGluThrArgGluThrValLeuGluSerGlyA 593
Db       1610 GTCTTACGGCGTACGTCACCCGCGATCCTGAAACCCGCCAGATGGTCCTCGAGTCGGGTG 1669

Qy        593 laLeuValLeuSerAspArgGlyIleCysCysIleAspGluPheAspLysMetSerAspA 613
Db       1670 CCTTGGTTCTTTCAGACGGCGGTATCTGTTGCATCGACGAGTTCGACAAGATGAACGAAT 1729

Qy        613 snAlaArgSerMetLeuHisGluValMetGluGlnGlnThrValSerValAlaLysGlyG 633
Db       1730 CCACTCGGTCCGTTCTGCATGAAGTCATGGAACAACAGACAGTATCTATCGCCAAGGCAG 1789

Qy        633 lyIleIleAlaSerLeuAsnAlaArgThrSerValLeuAlaCysAlaAsnProSerGlyS 653
Db       1790 GCATTATCACTACTTTGAACGCTAGGACCAGCATCCTGGCTTCCGCCAATCCGATCGGTA 1849

Qy        653 erArgTyrAsnAlaArgLeuSerValIleAspAsnIleGlnLeuProProThrLeuLeuS 673
Db       1850 GCAGGTACAATCCCAACTTGCCCGTTCCTCAAAATATTGACCTTCCGCCTACCTTGCTCT 1909

Qy        673 erArgPheAspLeuIleTyrLeuMetLeuAspLysProAspGluGlnAsnAspArgArgL 693
Db       1910 CCCGATTCGACTTGGTATACCTCGTGCTGGACCGAGTGGATGAGCAGGAAGATCGTCGGC 1969

Qy        693 euAlaArgHisLeuValAlaLeuHisTyrGluAsn------TyrGluValSerLysGlnA 711
Db       1970 TCGCTAAGCATCTTGTCAATATGTACCTGGAAGACAGACCTGAGCATGCTGCCGAGCAAG 2029

Qy        711 spAlaLeuAspLeuGlnThrLeuThrAlaTyrIleThrTyrAlaArgGlnHisValHisP 731
Db       2030 AAATCTTGCCGATCGAATTCCTTACAGCCTATATCACCTACGCCAAGACCAAAGTCCATC 2089

Qy        731 roThrLeuSerAspGluAlaAlaGluAspLeuIleAsnGlyTyrValGluMetArgGlnL 751
Db       2090 CAGTGCTCACACCGGCCGCCGGTAAAGCCTTGTCGGATGCTTACGTTAACATGCGTAAGC 2149

Qy        751 ysGlyAsnPheProGlySerSerLysLysValIleThrAlaThrProArgGlnLeuGluS 771
Db       2150 TTGGAGATGACATCCGGTCTTCTGACCGCCGTATCACCGCTACCACTCGTCAACTGGAGT 2209

Qy        771 erMetIleArgIleSerGluAlaLeuAlaArgMetArgPheSerGluValValGluLysV 791
Db       2210 CCATGATCCGACTGTCGGAAGCGCATGCGCGTATGCGGCTATCGCCGGAGGTCACTGCGG 2269

Qy        791 alAspAlaAlaGluAlaValArgLeuLeuAspValAlaLeuGlnGlnSerAlaThrAspH 811
Db       2270 ATGATGTGGAGGAAGCCGTGCGCCTGATCCGCTCCGCCATCAAGCAGGCGGCCACTGACT 2329

Qy        811 isAlaThrGlyThrIleAspMetAspLeuIleThrThrGlyValSerAlaSerGluArgI 831
Db       2330 CTCGGACCGGTCTGATCGACATGAGCTTGTTGACGGAGGGCACTAGTGCCAGCGAGAGAC 2389

Qy        831 leArgArgAlaAsnLeuLeuAlaAlaLeuArgGluLeuIleAlaAspLysIleSerProG 851
Db       2390 GCAGCCGGGAAGCACTCAAGCGTGCCTTGCTGTCTGTGGTGGATGATCTGTGCAGCGGTG 2449

Qy        851 lySerSerSerGlyLeuLysThrSerGlnLeuLeuGluAspIleArgSerGlnSerSerV 871
Db       2450 GCGGTGCAGCT------CGCTGGGCCGAGGTCTTCAGGATCTTAAGCGAGAACAGCAGCA 2503

Qy        871 alAspValSerLeuGlnAspIleLysAsnAlaLeuGlySerLeuGlnGlyGluGlyPheL 891
Db       2504 TTGAGGTGGATGGAGCCCAGTTTGCGGATGCGGTTCGAGCGCTGGAGGCTGAGGGAGCGG 2563

Qy        891 euThrValHisGlyAspIleValLysArg 900
Db       2564 TGAGTGTGGTCGGCGAGGGTGCGCGGCGG 2592
```

RESULT 6
ADJ39580

```
ID   ADJ39580 standard; cDNA; 1723 BP.
XX
AC   ADJ39580;
XX
DT   06-MAY-2004  (first entry)
XX
DE   Plant cDNA #580.
XX
KW   Plant; gene; ss; transcription; plant genome augmentation; cereal;
KW   soybean; alfalfa; sunflower; canola; cotton; peanut; tobacco; sugar beet;
KW   maize; barley; sorghum; rice; wheat; crop plant; insecticide resistance;
KW   stress tolerance; salt tolerance; cold tolerance; drought tolerance;
KW   plant nutrition; apical dominance; dwarfism; early flowering; antiviral;
KW   antifungal.
XX
OS   Eukaryota.
XX
PN   US2004016025-A1.
XX
PD   22-JAN-2004.
XX
PF   26-SEP-2002; 2002US-00260238.
XX
PR   26-SEP-2001; 2001US-0325277P.
PR   26-SEP-2001; 2001US-0325448P.
PR   04-APR-2002; 2002US-0370620P.
XX
PA   (BUDW/) BUDWORTH P.
PA   (MOUG/) MOUGHAMER T.
PA   (BRIG/) BRIGGS S P.
PA   (COOP/) COOPER B.
PA   (GLAZ/) GLAZEBROOK J.
PA   (GOFF/) GOFF S A.
PA   (KATA/) KATAGIRI F.
PA   (KREP/) KREPS J.
```

```
PA   (PROV/) PROVART N.
PA   (RICK/) RICKE D.
PA   (ZHUT/) ZHU T.
XX
PI   Budworth P,  Moughamer T,  Briggs SP,  Cooper B,  Glazebrook J;
PI   Goff SA,  Katagiri F,  Kreps J,  Provart N,  Ricke D,  Zhu T;
XX
DR   WPI; 2004-190374/18.
XX
PT   New rice promoter, useful for manipulating crop plants to alter or
PT   improve phenotypic characteristics, e.g. produce large quantities of oil
PT   or proteins, resistance to insecticides, virus or fungi, stress tolerance
PT   or high nutritional value.
XX
PS   Claim 70; SEQ ID NO 580; 230pp; English.
XX
CC   The invention relates to plant nucleotide sequences that direct seed-,
CC   leaf- and/or stem-, panicle-, root- or pollen-specific or -preferential
CC   or constitutive transcription of an operatively linked nucleic acid
CC   segment. The invention also relates to a method for augmenting a plant
CC   genome and a method of identifying a gene, where its expression is
CC   altered in the seed, leaf, stem, panicle, pollen, root or is constitutive
CC   in a plant cell. The plant is a cereal, e.g. soybean, alfalfa, sunflower,
CC   canola, cotton, peanut, tobacco or sugar beet, preferably maize, barley,
CC   sorghum, rice or wheat. The polynucleotides and the polypeptides they
CC   encode are useful for manipulating crop plants to alter or improve
CC   phenotypic characteristics, to produce large quantities of oil or
CC   proteins, to incur resistance to insecticides, viruses or fungi, and to
CC   incur stress tolerance (e.g. salt, cold or drought) to ensure the plants
CC   have a high nutritional value with reduced apical dominance or dwarfism,
CC   early flowering or altered metabolic pathways. This sequence represents a
CC   plant nucleic acid of the invention. Note: The sequence data for this
CC   patent did not form part of the printed specification but was obtained in
CC   electronic format directly from USPTO at seqdata.uspto.gov/sequence.html.
XX
SQ   Sequence 1723 BP; 459 A; 373 C; 461 G; 429 T; 0 U; 1 Other;

Alignment Scores:
Pred. No.:              1.42e-99     Length:          1723
Score:                  1840.00      Matches:         403
Percent Similarity:     64.41%       Conservative:    82
Best Local Similarity:  53.52%       Mismatches:      74
Query Match:            39.92%       Indels:          196
DB:                     12           Gaps:            12

US-10-768-511-8 (1-901) x ADJ39580 (1-1723)

Qy        135 LeuHisArgArgArgArgSerGlnSerArgGluProGlyHisArgSerProSerArgGlu 154
                :::||||||||||||              |||    ||||||    |||:::|||
Db          3 GTCCACCGACGACGCCC-GGTTCTCCTCGGAGGCCGGGGACGAGGACACCC--------- 52

Qy        155 ProSerAlaAspGlyArgProSerGluSerAlaGluProAspAspThrLeuGlyGlyGlu 174
              |||      |||||                |||      |||   |||
Db         53 CCGAG-ACCGACGGCGGCGGCGGCGGCGGCGCGGGCGCCGACGNCACG------------ 99

Qy        175 TyrAlaTyrValTrpGlyThrAsnValAsnIleProAspValLeuArgAlaIleArgArg 194
                       :::||||||||||||||||:::::::::     ||||||      ||||||    |||
Db        100 CCGGTGTTCGTCTGGGGCACCAACATCAGCGTGCAGGACGTCAACGCCGCCATCCTCCGG 159

Qy        195 PheLeuHisAsnTyrArgSerSerAlaHisAspLeuAsnSerLysTyrIleGlnIleIle 214
              ||||||    ::::::|||:::     |||
Db        160 TTCCTCCGCCACTTCCGGACCCGCGCG--------------------------------- 186

Qy        215 GluGluThrValGluArgGluGluAspThrLeuAsnIleAspMetSerAspIleTyrAsp 234

Db        186 ------------------------------------------------------------ 186

Qy        235 HisAspProAspLeuTyrAlaLysIleValArgTyrProLeuAspIleIleProLeuLeu 254
                                                                |||
Db        187 ------------------------------------------------ACGCCGGCCGCG 198

Qy        255 AspThrGluCysGlnGluValAlaThrSerLeuLeuProThrPheGluLysHisIleGlu 274
                 |||

Db        199 TCGACC------------------------------------------------------ 204

Qy        275 AlaArgProPheAsnLeuLysAlaSerValHisMetArgGluLeuAsnProSer-AspIl 294
                                                  |||               ::: |||||
Db        205 ----------------------------------CGGTCATGGACGAGGGCAAGATAT 228

Qy        294 eAspLysLeuValSerValLysGlyMetValIleArgCysSerSerIleIleProGluIl 314
              |:::|||:::||||||:::|||||||||:::||||||||||||||||:::|||||||||::
Db        229 CGAGAAGATGGTGTCCATCAAGGGTATGATAATTCGGTGCAGCTCGGTGATTCCAGAGCT 288

Qy        314 eLysGlyAlaPhePheLysCysLeuValCysGly---HisSerProProLeuValThr-V 333
              :|||    |||   |||:::||||||||||||||     :::|||    |||:::::     :
Db        289 CAAGGAGGCTGTGTTCCGCTGCTTGGTTTGCGGGTTCTACTCTGAGCCTGTAATGGTTGA 348

Qy        333 alValLysGlyArgValGluGluProThrArgCysGluLysProGluCysAlaAlaArgA 353
              ::                                                          |
Db        349 TC----------------------------------------------------GAGA 354

Qy        353 snAlaMetSerLeuIleHisAsnArgCys--ThrPheAlaAsnLysGlnIleValArgLe 372
              ||:::|||:::|||:::||||||||||||      ||||||:::||||||||||::::::||
Db        355 ACTCTATGACTCTTGTGCATAACCGATGCAGAGGTTTGCAGATAAGCAGATCATAAAGTT 414

Qy        372 uGlnGluThrProAspAlaIleProGluGlyGluThrProHisThrValSerMetCysLe 392
              |||||||||||||||    ||||||||||||    |||||||||||||||||||:::    ::
Db        415 GCAGGAAACACCAGATGAGATACCAGAAGGTGGCACTCCACATACTGTTAGTGTCTTGAT 474

Qy        392 uTyrAsnThrMetValAspAlaValLysProGlyAspArgIleGluValThrGlyValPh 412
              :::::::    :::||||||||    ||||||||||||:::::|||    ||||||::::::
Db        475 GCACGATAAGCTTGTAGATGCTGGAAAGCCTGGAGACAGGGTTGAG---ACTGGGATATA 531

Qy        412 eLysAlaMetAlaValArgValGlyProAsnGlnArgThrLeuArgAlaLeuTyrLysTh 432
              ::::||||||::::::||||||||||||    |||||||||:::::::::::::::|||||
Db        532 CAGAGCCATGAGTATTAGAGTTGGCCCAACTCAGAGGACAGTGAAGTCGATTTTCAAGAC 591

Qy        432 rTyrIleAspCysValHisValLysLysSerAspArgGlyArgLeuGlnThrGluAspPr 452
              ||||||||||||:::|||:::|||
Db        592 GTACATTGATTGCCTTCACATAAAG----------------------------------- 616

Qy        452 oMetGluMetAspLysGluAsnAspMetTyrAlaGlyTyrHisGluSerAspThrSerGl 472

Db        616 ------------------------------------------------------------ 616

Qy        472 uAlaAlaAsnGluAlaLysIleGlnLysLeuLysGluLeuSerLysLeuProGlyIleTy 492
                                          ||||||||||||||||||||    |||||
Db        617 ------------------------------AAGGAGTTGTCAAAGTTGCCAGATATATA 645

Qy        492 rAspArgLeuSerArgSerLeuAlaProSerIleTrpGluLeuGluAspIleLysLysGl 512
              |||||||||:::|||||||||||||||||||||||||||||:::|||:::||||:::|||
Db        646 TGACAGATTAACTAGGTCATTGGCTCCAAACATATGGGAGCTGGACGACGTTAAAAGGGG 705

Qy        512 yLeuLeuCys---GlnLeuPheGlyGlyLysAlaLysLysIleProSerGlyAlaSerPh 531
              |||||||||     ||||||||||||||    |||     ::::::||||||||||||||||
Db        706 CCTCCTTTGCCAGCAGCTTTTTGGTGGAAATGCTTTGAGGCTTCCTTCTGGAGCTAGTTT 765

Qy        531 eArgGlyAspIleAsnValLeuLeuValGlyAspProGlyThrSerLysSerGlnLeuLe 551
              ||||||||||||||:::||||||||||||||||||||||||||||||||||||||||||||
Db        766 CCGAGGCGACATCAATATTTTGCTTGTTGGTGATCCTGGAACAAGTAAATCCCAGCTTCT 825

Qy        551 uGlnTyrValHisLysIleAlaProArgGlyIleTyrThrSerGlyArgGlySerSerAl 571
              ||||||:::||||||::::::||||||||||||||||||||||||||||||||||||||||
Db        826 CCAATACATGCACAAACTGTCTCCTCGTGGCATTTATACAAGTGGCAGAGGAAGTTCAGC 885

Qy        571 aValGlyLeuThrAlaTyrValThrLysAspProGluThrArgGluThrValLeuGluSe 591
              ||||||||||||||||||||||||||||||||||||||||    ||||||||||||||||||
Db        886 TGTTGGCCTTACTGCATACGTTACCAAGGATCCTGAAACTGGTGAAACTGTTCTTGAGAG 945

Qy        591 rGlyAlaLeuValLeuSerAspArgGlyIleCysCysIleAspGluPheAspLysMetSe 611
              ||||||||||||||||||||||||:::|||:::|||||||||||||||||||||||||||||
Db        946 TGGAGCGCTTGTTTTGAGTGACAAAGGTGTTTGTTGTATTGATGAATTTGATAAGATGTC 1005

Qy        611 rAspAsnAlaArgSerMetLeuHisGluValMetGluGlnGlnThrValSerValAlaLy 631
              |||||||||||||||||||||||||||||||||||||||||||||||||||:::||||||
Db       1006 TGATAATGCTCGAAGCATGTTACATGAGGTGATGGAACAACAGACTGTCTCCATTGCCAA 1065
```

```
Qy        631 sGlyGlyIleIleAlaSerLeuAsnAlaArgThrSerValLeuAlaCysAlaAsnProSe 651
              |  ||||||||||||||||||||||||||||||||||||||||||||||||||||||:::
Db       1066 GGCTGGAATAATTGCATCTTTAAATGCCAGAACATCAGTTCTAGCATGTGCAAATCCTAC 1125

Qy        651 rGlySerArgTyrAsnAlaArgLeuSerValIleAspAsnIleGlnLeuProProThrLe 671
              :   ||||||||||||   ||||||||||||||||||||||||    ||||||||||||||
Db       1126 TGAATCACGTTATAATCCAAGGCTCTCTGTGATTGACAATATCCATCTTCCTCCAACACT 1185

Qy        671 uLeuSer--ArgPheAspLeuIleTyrLeuMetLeuAspLysProAspGluGlnAsnAsp 690
              ||||||  |||||||||||||||||||||:::||||||||   ||||||||    |||
Db       1186 GCTTTCTAGAGGTTTGACCTCATTTATCTGATATTGGACAAGGCAGATGAGCAAACTGAT 1245

Qy        691 ArgArgLeuAlaArgHisLeuValAlaLeuHisTyrGluAsnTyrGluValSerLysGln 710
              ||||||||||:::|||:::|||:::||||||:::||||||
Db       1246 AGACGCCTGGCTAAGCATATTGTTTCGTTGCATTTTGAGAAT------------------ 1287

Qy        711 AspAlaLeuAspLeuGlnThrLeuThrAlaTyrIleThrTyrAlaArgGlnHisValHis 730
                                                 ||||||:::||||||||:::|||:::
Db       1288 ------------------------CCAAACTACATAAGTTATGCAAGGAAGCATATACAA 1323

Qy        731 ProThrLeuSerAspGluAlaAlaGluAspLeuIleAsnGlyTyrValGluMetArgGln 750
              |||  |||||||||||||||||||||||:::|||       |||||||||||||||||:::
Db       1324 CCACAGTTATCTGATGAAGCTGCAGAAGAATTGACTCGCGGCTATGTTGAGATGAGGAAA 1383

Qy        751 LysGlyAsnPheProGlySerSerLysLysValIleThrAlaThrProArgGlnLeuGlu 770
              :::||||||  |||||||||  ||||||  ||||||||||||  ||||||:::|||
Db       1384 AGAGGAAACAGCCCTGGTAGCAGAAAGAAG---ATAACTGCGACAGCTCGACAAATTGAG 1440

Qy        771 SerMetIleArgIleSerGluAlaLeuAlaArgMetArgPheSerGluValValGluLys 790
              |||:::||||||:::||||||||||||||||||||||||||||||||||:::||||||
Db       1441 AGCTTGATTCGGCTCAGTGAAGCACTGGCCCGAATGCGATTCTCTGAAATGGTTGAAGTA 1500

Qy        791 ValAspAlaAlaGluAlaValArgLeuLeuAspValAlaLeuGlnGlnSerAlaThrAsp 810
                 |||      ||||||   |||||||||:::||||||:::||||||||||||||||||
Db       1501 CAAGATGTTGTAGAGGCCTTCAGGCTTCTCGAAGTTGCCATGCAGCAATCGGCAACTGAT 1560

Qy        811 HisAlaThr-GlyThrIleAspMetAspLeuIleThrThrGlyValSerAlaSerGluAr 830
              ||||||||| ||||||||||||||||||||||||     ||||||:::||||||||||||||
Db       1561 CATGCCACTGGGTACAATCGATATGGATCTTATCATGACTGGAATATCTGCGAGCGAAAG 1620

Qy        830 gIleArgArgAlaAsnLeuLeuAlaAlaLeuArgGluLeuIleAlaAspLysIleSerPr 850
              |  ||||||
Db       1621 GCAGAGGCGG-------------------------------------------------- 1630

Qy        850 oGlySerSerSerGlyLeuLysThrSerGlnLeuLeuGluAspIleArgSerGlnSerSe 870
                                        ||||||||||||:::||||||   ||||||||
Db       1631 ------------------------GACCAGTTGCTGGAAGAAATTAGGAAGCAGAGCTC 1665

Qy        870 rValAspValSerLeuGlnAspIle 878
              |::::::|||   |||   |||:::
Db       1666 TATGGAAGTTCATCTGCATGATGTA 1690

RESULT 7
ADB69927
ID   ADB69927 standard; DNA; 2970 BP.
XX
AC   ADB69927;
XX
DT   04-DEC-2003  (first entry)
XX
DE   C. neoformans open reading frame SEQ ID NO:2332.
XX
KW   ds; gene; fungicide; gene therapy; infection.
XX
OS   Cryptococcus neoformans.
XX
PN   WO2003052076-A2.
XX
PD   26-JUN-2003.
XX
PF   17-DEC-2002; 2002WO-US040225.
XX
PR   17-DEC-2001; 2001US-0341261P.
XX
PA   (ELIT-) ELITRA PHARM INC.
XX
PI   Zamudio C,  Eroshkin AM;
XX
DR   WPI; 2003-533017/50.
DR   P-PSDB; ADB70288.
XX
PT   New nucleic acid, useful for preparing a composition for treating an
PT   infection caused by Cryptococcus neoformans.
XX
PS   Claim 2; SEQ ID NO 2332; 136pp; English.
XX
CC   The invention relates to a novel purified or isolated Cryptococcus
CC   neoformans nucleic acid molecule comprising a sequence encoding a
CC   polypeptide comprising a sequence not given in the specification. A
CC   polynucleotide of the invention has fungicide activity, and may have a
CC   use in gene therapy. The nucleic acid is useful for preparing a
CC   composition for treating an infection caused by Cryptococcus neoformans.
CC   The present sequence represents a C. neoformans sequence of the
CC   invention. Note: The sequence data for this patent is not represented in
CC   the printed specification, but was obtained in electronic format directly
CC   from WIPO at ftp.wipo.int/pub/published_pat_sequences.
XX
SQ   Sequence 2970 BP; 711 A; 746 C; 742 G; 771 T; 0 U; 0 Other;

Alignment Scores:
Pred. No.:              4.93e-99      Length:         2970
Score:                  1835.50       Matches:        409
Percent Similarity:     59.59%        Conservative:   147
Best Local Similarity:  43.84%        Mismatches:     246
Query Match:            39.82%        Indels:         131
DB:                     10            Gaps:           20

US-10-768-511-8 (1-901) x ADB69927 (1-2970)

Qy         47 AlaValArgGlnThrProThrSerAlaValArgArg--------------ArgGlyArg 61
              |||:::   |||:::      ||||||::::::                  ||||||
Db         10 GCTCTGCTGCAGTCAGACGAATCCGCGATGAAACACCATTGTTCTTCCCTCCGTGGTAGT 69

Qy         62 GluThrAspSerAlaArgArgArgArgSerArgSerArgSerLeuGlyAsnSerValTyr 81
                       |||     ||||||     |||                 :::    :::
Db         70 TCCGGAGGATCGACCCCTCGTCGAGCTCGTCGAGGTGACATCCACTCCTCTTTCCCGCAC 129

Qy         82 SerSerPro------------------------------------------TyrAsp 86
              |||||||||
Db        130 TCATCGCCGTCTCTCGCTCGACGACAAGCCCAAGTCCCGGCTGACTCTCTTTCGCTCCGT 189

Qy         87 AlaGlyThrProGlyThrProGlyThrProValAlaThrProValTyrAlaThrProVal 106
              |||   :::||||||
Db        190 GCGTCTTCCCCTGGT--------------------------------------------- 204

Qy        107 GlyThrProMetGlyThrProSerPheHisArgGlyThrProGlnTyrLysGlnArgSer 126
                       |||||||||   ||||||      ||||||||||:::
Db        205 ---------ATGGGTACCGATAGCTTT------GGTACTCCGCGTGTT----------- 237

Qy        127 GluLeuGlySerGlnGlyLysProLeuHisArgArgArgSerGlnSerArgGluPro 146
                    ||||||         |||          :::||||||
Db        238 ---TTCGGATCGGCTGCCCCTACCCTTTCCGCAGTTGCCCAATCTCAACTT--------- 285

Qy        147 GlyHisArgSerProSerArgGluProSerAlaAspGlyArgProSerGluSerAlaGlu 166
                                           |||||||||
Db        286 --------------------------GGTGCCGACGGC-------------------- 297

Qy        167 ProAspAspThrLeuGlyGlyGluTyrAlaTyrValTrpGlyThrAsnValAsnIlePro 186
                    |||   :::   |||         ::::::|||||||||   :::::::::
Db        298 ------GACGATATCGATGGTATGGTCAAATTTATCTGGGGTACGACAATCTCTTTGCAA 351

Qy        187 AspValLeuArgAlaIleArgArgPheLeu----------------------------- 196
              :::   :::         |||   ||||||
Db        352 GAGAGCATGAACCTTTTCCGGGATTTTTTGAGAGGCTTCAAACCCAAGTACCGTGCTGTG 411
```

```
Qy    197 HisAsnTyrArgSerSerAlaHisAspLeuAsnSer------------------------ 208
Db    412 TACAATGCGGAACAATCCAGACATGCTGAGGAGAGTGGCGGCGTTGCGCCTCCTCCGATG 471

Qy    209 ---LysTyrIleGlnIleIleGluGluThrValGlu------------------------ 219
Db    472 ACTTTGTACGACAATCTCTCTGCTGAGCGTGCCGAGGTCCCTCTCTATGAGACGTATCTC 531

Qy    220 ---------------ArgGluGluAspThrLeuAsnIleAspMetSerAspIleTyrAsp 234
Db    532 AACCGGCTTCGACTCACTGGTGAAACTAATTTAAACCTCGACGCGCTCAATCTGCTCGCG 591

Qy    235 HisAspPro------AspLeuTyrAlaLysIleValArgTyrProLeuAspIleIlePro 252
Db    592 TACCGACCCACAAAAAAGCTTTACCAGCAATTGGTCAACTATCCGCAAGAAGTGATCCCT 651

Qy    253 LeuLeuAspThrGluCysGlnGluValAlaThr--------------------------- 263
Db    652 ATCATGGATCAGGTCTTGAGGGATGTGATGATTGAATTGGGTCATGAAGAGCTGGAGAAA 711

Qy    264 -----------------------SerLeuLeuProThrPheGluLysHisIle---Glu 274
Db    712 GCGCAGACCAAATTCGCTGAGGGTAACTTATCTCAAGATGTGGAGTCACGAGTGTATAAA 771

Qy    275 AlaArgProPheAsnLeuLysAlaSerValHisMetArgGluLeuAsnProSerAspIle 294
Db    772 GTCAGACCGTTTGGTGGAGAGAAGACGGTCAACATGAGGGATCTTAATCCCGGAGATACT 831

Qy    295 AspLysLeuValSerValLysGlyMetValIleArgCysSerSerIleIleProGluIle 314
Db    832 GACAAGCTGGTGACCGTGAAGGGTCTTGTCATCCGTGCAACTCCCGTCATTCCTGATATG 891

Qy    315 LysGlyAlaPhePheLysCysLeuValCysGlyHisSerProProLeuValThrValVal 334
Db    892 ACCACTGCTTTCTTCCGATGCTTGGTTTGTCAGCACACTGTTCAA---GCGGATATCGAC 948

Qy    335 LysGlyArgValGluGluProThrArgCysGluLysProGluCysAlaAlaArgAsnAla 354
Db    949 CGAGGAAGAATCAGCGAGCCAGAGCGATGTCCCCGAGACGTCTGTGGTTCAACTGGCACT 1008

Qy    355 MetSerLeuIleHisAsnArgCysThrPheAlaAsnLysGlnIleValArgLeuGlnGlu 374
Db   1009 ATGTCTCTTATTCATAACCGCTCCGAATTCACCGACAAACAAGTCATCCGTCTTCAGGAG 1068

Qy    375 ThrProAspAlaIleProGluGlyGluThrProHisThrValSerMetCysLeuTyrAsn 394
Db   1069 ACTCCCGATGCTGTACCCGATGGGCAGACTCCACATACTGTATCTCTTTGCGTCTATGAT 1128

Qy    395 ThrMetValAspAlaValLysProGlyAspArgIleGluValThrGlyValPheLysAla 414
Db   1129 GAACTCGTTGACTTGGTCAAGCCTGGTGACCGAGTGATCATCACTGGTATCTTCCGATCA 1188

Qy    415 MetAlaValArgValGlyProAsnGlnArgThrLeuArgAlaLeuTyrLysThrTyrIle 434
Db   1189 ATCCCAGTTCGTGTCAACCCTCGCCAAAGGAGTATCAAATCTTTGTACAAGACCTATCTC 1248

Qy    435 AspCysValHisValLysLysSerAspArgGlyArgLeuGlnThrGluAspProMetGlu 454
Db   1249 GATGTCGTCCACGTCAAGCGTACCAACACTGCCCGAATGGCTTT---GACCCTTCTACT 1305

Qy    455 MetAspLysGluAsnAspMetTyrAlaGlyTyrHisGluSerAspThrSerGluAlaAla 474
Db   1306 CGTGCGGGCGAGAGCAAGCCA---CCGGGTGACCTTGGGTCGCCTGTCAGATCAGCCGCT 1362

Qy    475 AsnGluAlaLysIleGlnLysLeuLysGluLeuSerLysLeuProGlyIleTyrAspArg 494
Db   1363 GCGGAGATGGAA---CAACGCATTATCGAGCTCTCCAACCACCCTGATCTTTACAACATA 1419

Qy    495 LeuSerArgSerLeuAlaProSerIleTrpGluLeuGluAspIleLysLysGlyLeuLeu 514
Db   1420 CTTGCCTCATCTTTGGCGCCCTCCATCTACGAGCTCGAAGATGTTAAGAAGGGCATCCTC 1479

Qy    515 CysGlnLeuPheGlyGlyLysAlaLysLysIleProSerGlyAlaSer------------ 530
Db   1480 CTCCAGCTTTTCGGCGGTACCAACAAGTCTATCGCCCGTGGTGGCGGTGGGGGCGGTCCT 1539

Qy    531 ---PheArgGlyAspIleAsnValLeuLeuValGlyAspProGlyThrSerLysSerGln 549
Db   1540 AGATACAGAGGTGATATCAACGTGTTGATGGTTGGTGATCCTGGTACTAGTAAATCTCAG 1599

Qy    550 LeuLeuGlnTyrValHisLysIleAlaProArgGlyIleTyrThrSerGlyArgGlySer 569
Db   1600 ATTCTGCAGTATGTGCATAAGATTGCTCCTCGAGGTGTCTACACCTCCGGTAAGGGAAGT 1659

Qy    570 SerAlaValGlyLeuThrAlaTyrValThrLysAspProGluThrArgGluThrValLeu 589
Db   1660 TCAGCGGTTGGTTTGACGGCCTATGTAACGAGAGATCCCGATTCCAAGCAGCTCGTTTTG 1719

Qy    590 GluSerGlyAlaLeuValLeuSerAspArgGlyIleCysCysIleAspGluPheAspLys 609
Db   1720 GAAAGTGGTGCATTGGTTTTGTCTGACGGCGGAGTTTGCTGTATCGATGAGTTTGACAAG 1779

Qy    610 MetSerAspAsnAlaArgSerMetLeuHisGluValMetGluGlnGlnThrValSerVal 629
Db   1780 ATGAGTGATGCCACTCGAAGTGTCTTGCACGAAGTAATGGAGCAACAAACGGTTTCTATT 1839

Qy    630 AlaLysGlyGlyIleIleAlaSerLeuAsnAlaArgThrSerValLeuAlaCysAlaAsn 649
Db   1840 GCCAAGGCAGGTATCATCACCACTCTCAATGCTCGTACATCCATTCTCGCCGCCGCGAAC 1899

Qy    650 ProSerGlySerArgTyrAsnAlaArgLeuSerValIleAspAsnIleGlnLeuProPro 669
Db   1900 CCAATCAACTCTCGATATGACCCTAACCTCCCTATTCCTGCCAACATCGATCTCCCTCCT 1959

Qy    670 ThrLeuLeuSerArgPheAspLeuIleTyrLeuMetLeuAspLysProAspGluGlnAsn 689
Db   1960 ACATTGATCTCAAGATTCGACTTGCTGTACTTGGTGTTGGACAAGGTGGATGAGGTGAAC 2019

Qy    690 AspArgArgLeuAlaArgHisLeuValAlaLeuHisTyrGluAsnTyrGluValSerLys 709
Db   2020 GACAGAAAGTTGGCAAAGCATTTGGTGGGACTCTATTTGTCAGATGTAGAGGATCAGCCT 2079

Qy    710 GlnAsp---AlaLeuAspLeuGlnThrLeuThrAlaTyrIleThrTyrAlaArgGlnHis 728
Db   2080 GCCGACAATATCATTCCTCTTCAGACCTTAACGTCCTACATTACATATGCCCGCTCCAAA 2139

Qy    729 ValHisProThrLeuSerAspGluAlaAlaGluAspLeuIleAsnGlyTyrValGluMet 748
Db   2140 ATCCACCCCGTCCTCACCGAAGGAGCTTCTGAAGCGCTTGTTCAGGCCTATGTTGAGATG 2199

Qy    749 ArgGlnLysGlyAsnPheProGlySerSerLysLysValIleThrAlaThrProArgGln 768
Db   2200 CGCAAGGCCGGTATGGACTCTCGTACTCAGGAAAAGAGGATCACTGCTACCACTAGGCAG 2259

Qy    769 LeuGluSerMetIleArgIleSerGluAlaLeuAlaArgMetArgPheSerGluValVal 788
Db   2260 TTGGAAAGTATGATCAGATTGGGCGAAGCGCATGCGAGGATGAGATTAAGTGACAGGGTT 2319

Qy    789 GluLysValAspAlaAlaGluAlaValArgLeuLeuAspValAlaLeuGlnGlnSerAla 808
Db   2320 GAGGAAGAGGACATTAGGGAGGCTGTTAGGTTGATCAAGAGTGCCCTGAGGGAGAGTGCT 2379

Qy    809 ThrAspHisAlaThrGlyThrIleAspMetAspLeuIleThrThrGlyValSerAlaSer 828
Db   2380 ACCGATCCCCTTACTGGTCAAATTGATCTTGACCTCATCAACACCGGTGCCGGCCAAACC 2439

Qy    829 GluArgIleArgArgAlaAsnLeuLeuAlaAlaLeuArgGluLeuIleAlaAspLysIle 848
Db   2440 ATGCGTCGCGCCCGAGCCGATCTTAAACGTGAAGTTATCAAACTCGTCGTCGAAAAG--- 2496

Qy    849 SerProGlySerSerSerGlyLeuLysThrSerGlnLeuLeuGluAspIleArgSerGln 868
Db   2497 ------GCCCGATCTCAGGGTATTCGCTGGGCAGCAGTCATTGATGAGCTCAACAAACAA 2550

Qy    869 SerSerValAspValSerLeuGlnAspIleLysAsnAlaLeuGlySerLeuGlnGlyGlu 888
```

```
Db       2551 TCCAGTGTTCCCGTGGATCATGCACAGTTCGCAGAAATTGTTAGGGAGCTGGAAGAAGAG 2610

Qy        889 GlyPheLeuThrValHisGlyAspIleValLysArgVal 901
                 :::   |||   |||:::      :::|||:::
Db       2611 AGCATAGTCAAGGTCATGGGTGAAAGGGAAAGAAGGATT 2649

RESULT 8
ADT47760
ID   ADT47760 standard; cDNA; 2802 BP.
XX
AC   ADT47760;
XX
DT   02-DEC-2004  (first entry)
XX
DE   Bacterial polynucleotide #22511.
XX
KW   Recombinant DNA construct; transformed plant; improved plant property;
KW   cold tolerance; heat tolerance; drought tolerance; herbicide; osmosis;
KW   pathogen tolerance; pest tolerance; plant disease resistance;
KW   cell cycle pathway modification; plant growth regulator;
KW   homologous recombination; seed oil yield; protein yield; carbohydrate;
KW   nitrogen; phosphorus; photosynthesis; lignin; galactomannan;
KW   bacterial polynucleotide; gene; ss.
XX
OS   Bacteria.
XX
PN   US2003233675-A1.
XX
PD   18-DEC-2003.
XX
PF   20-FEB-2003; 2003US-00369493.
XX
PR   21-FEB-2002; 2002US-0360039P.
XX
PA   (CAOY/) CAO Y.
PA   (HINK/) HINKLE G J.
PA   (SLAT/) SLATER S C.
PA   (CHEN/) CHEN X.
PA   (GOLD/) GOLDMAN B S.
XX
PI   Cao Y,  Hinkle GJ,  Slater SC,  Chen X,  Goldman BS;
XX
DR   WPI; 2004-061375/06.
XX
PT   New recombinant DNA construct comprising a promoter positioned to provide
PT   for expression of a polynucleotide encoding a polypeptide from a
PT   microbial source, useful for producing plants with improved properties.
XX
PS   Claim 1; SEQ ID NO 46198; 122pp; English.
XX
CC   The invention relates to a recombinant DNA construct comprising a
CC   promoter functional in a plant cell, where the promoter is positioned to
CC   provide for expression of a polynucleotide encoding a polypeptide from a
CC   microbial source. The invention also relates to a transformed plant
CC   comprising the recombinant DNA construct and a method of producing a
CC   transformed plant having an improved property. The plant is a crop plant
CC   such as maize or soybean. The method of producing a transformed plant
CC   having an improved property comprises transforming a plant with the
CC   recombinant DNA construct and growing the transformed plant, where the
CC   polynucleotide or polypeptide is useful for improving plant properties.
CC   The recombinant DNA construct is useful for producing plants with
CC   improved plant properties, e.g. improved cold, heat or drought tolerance,
CC   tolerance to herbicides, extreme osmotic conditions, pathogens or pests,
CC   increased resistance to plant disease, better growth rate by modification
CC   of the cell cycle pathway with plant growth regulators, increased rate of
CC   homologous recombination, modified seed oil or protein yield and/or
CC   content, improved yield by modification of carbohydrate, nitrogen or
CC   phosphorus use and/or uptake, by modification of photosynthesis or by
CC   providing improved plant growth and development under at least one stress
CC   condition, improved lignin production or improved galactomannan
CC   production. This sequence represents a bacterial polynucleotide used in
CC   the scope of the invention. Note: The sequence data for this patent did
CC   not form part of the printed specification but was obtained in electronic
CC   format from USPTO at seqdata.uspto.gov/sequence.html.
XX
SQ   Sequence 2802 BP; 887 A; 541 C; 579 G; 795 T; 0 U; 0 Other;

Alignment Scores:
Pred. No.:               6.48e-99      Length:        2802
Score:                   1833.00       Matches:       393
Percent Similarity:      59.36%        Conservative:  178
Best Local Similarity:   40.85%        Mismatches:    281
Query Match:             39.77%        Indels:        110
DB:                      13            Gaps:          14

US-10-768-511-8 (1-901) x ADT47760 (1-2802)

Qy          2 GluAsnAsnAspAlaLeuAspIleGlyAlaValSerSerProTyrProSerGlnSerGlu 21
              |||:::|||::::::                  |||||||||               :::
Db         31 GAGGATAATAACTCC------------------AGCTCCCCCGTTGTGCCTAATCCTGAT 72

Qy         22 GlyValSerThrProLeuProGlnValThrSerProSerPheAspAsnAlaAlaSerPro 41
                 |||         ||||||::::::||||||:::          ::::::|||
Db         73 TCTGTTCCA---------CCACAGCTTTCTTCCCCAGCTCTATTTTATAGCTCCTCTTCA 123

Qy         42 ValAlaGly--------------------------------------------------- 44
                    |||
Db        124 TCACAAGGTGATATCTATGGTCGCAACAATAGCCAGAACTTAAGTCAGGGAGAGGGAAAC 183

Qy         45 ---ArgArgAlaValArgGlnThrPro---------ThrSerAlaValArgArgArgGly 60
                 |||   |||:::   :::|||            |||:::   :::|||:::
Db        184 ATCAGAGCTGCTATAGGTTCTTCTCCACTAAATTTTCCATCTTCTTCCCAAAGACAAAAT 243

Qy         61 ArgGluThrAspSerAlaArgArgArgArgSerArgSerArgSerLeuGlyAsnSerVal 80
                 :::      ::::::   |||:::   |||   ||||||       ::::::
Db        244 TCCGATGTTTTCCAATCTCAAGGCAGACAGGGCAGAATTCGTTCTTCTGCGAGCGCTTCT 303

Qy         81 TyrSerSerProTyrAspAlaGlyThrProGlyThrProGlyThrProValAlaThrPro 100
                       |||   |||   :::                   |||   ::::::
Db        304 GGAAGGTCTAGATATCATTCTGATCTGAGAAGTGATAGAGCACTGCCTACTTCTTCC--- 360

Qy        101 ValTyrAlaThrProValGlyThrProMetGlyThrProSerPheHisArgGlyThrPro 120

Db        360 ------------------------------------------------------------ 360

Qy        121 GlnTyrLysGlnArgSerGluLeuGlySerGlnGlyLysProLeuHisArgArgArgArg 140
                             |||   ||||||      |||:::                ||||||
Db        361 ---------------TCTTCTTTAGGCCGTAATGGTCAAAACCGTGTACACATGCGAAGA 405

Qy        141 SerGlnSerArgGluProGlyHisArgSerProSerArgGluProSerAlaAspGlyArg 160
              :::                           ||||||   |||            |||   |||
Db        406 AATGATATTCATACATCTGATTTATCCTCTCCAAGAAGAATTGTGGATTTTGATACTAGA 465

Qy        161 ------------------------------------------ProSerGluSerAlaGlu 166
                                                        |||||||||::::::|||
Db        466 TCCGGCGTGAATACTTTGGATACTTCTTCTTCTTCTGCTCCTCCATCGGAAGCCAGTGAA 525

Qy        167 ProAspAspThrLeuGlyGlyGluTyrAlaTyrValTrpGlyThrAsnValAsnIlePro 186
              |||                           :::||||||||||||||||:::|||
Db        526 CCC---------------------TTGAGAATAATTTGGGGTACCAACGTCAGTATCCAG 564

Qy        187 AspValLeuArgAlaIleArgArgPheLeuHisAsnTyrArg------------------ 200
              :::                  |||   ||||||   :::::::::
Db        565 GAATGTACAACTAATTTTCGTAATTTTTTGATGTCTTTCAAGTATAAATTCCGTAAAATA 624

Qy        201 ------------------------SerSerAlaHisAspLeuAsnSerLysTyrIleGln 212
                                      ::::::      |||            ||||||:::
Db        625 TTGGATGAAAGGGAGGAATTCATTAACAATACTACCGACGAAGAACTATACTATATCAAG 684

Qy        213 IleIleGluGluThrValGluArgGluGluAspThrLeuAsnIleAspMetSerAspIle 232
                 :::   |||      |||               ||||||:::|||         ::::::
Db        685 CAGCTTAATGAAATGAGAGAACTTGGTACCTCTAATTTAAACTTGGATGCTAGAAACTTA 744

Qy        233 ------TyrAspHisAspProAspLeuTyrAlaLysIleValArgTyrProLeuAspIle 250
                    |||         |||||||||   :::::::::   ||||||   ::::::
Db        745 CTTGCTTACAAGCAAACAGAAGACTTATACCATCAACTATTAAATTATCCTCAAGAAGTG 804
```

```
Qy        251 IleProLeuLeuAspThrGluCysGlnGluValAlaThrSerLeu--------------- 265
Db        805 ATTTCTATCATGGATCAAACTATCAAGGACTGTATGGTTTCCTTAATAGTAGACAACAAT 864
Qy        266 ------------LeuProThrPheGlu---LysHisIleGluAlaArgProPheAsnLeu 280
Db        865 TTGGATTACGATTTAGATGAAATAGAGACCAAGTTTATAAAGTAAGGCCTTACAATGTA 924
Qy        281 LysAlaSerValHisMetArgGluLeuAsnProSerAspIleAspLysLeuValSerVal 300
Db        925 GGTTCTTGTAAAGGTATGCGCGAATTGAATCCAAATGATATTGATAAATTGATAAACTTA 984
Qy        301 LysGlyMetValIleArgCysSerSerIleIleProGluIleLysGlyAlaPhePheLys 320
Db        985 AAGGGTCTTGTCCTGAGATCAACTCCGGTTATCCCTGATATGAAGGTAGCGTTTTTCAAA 1044
Qy        321 CysLeuValCysGlyHisSerProProLeuValThrValValLysGlyArgValGluGlu 340
Db       1045 TGCAACGTCTGCGATCATACA---ATGGCAGTGGAAATTGATAGAGGAGTTATACAAGAG 1101
Qy        341 ProThrArgCysGluLysProGluCysAlaAlaArgAsnAlaMetSerLeuIleHisAsn 360
Db       1102 CCCGCTAGGTGTGAACGTATTGATTGTAATGAACCCAATTCCATGTCACTGATTCACAAC 1161
Qy        361 ArgCysThrPheAlaAsnLysGlnIleValArgLeuGlnGluThrProAspAlaIlePro 380
Db       1162 AGGTGTTCATTCGCAGATAAACAAGTCATTAAGTTACAGGAAACTCCAGACTTTGTGCCT 1221
Qy        381 GluGlyGluThrProHisThrValSerMetCysLeuTyrAsnThrMetValAspAlaVal 400
Db       1222 GATGGACAAACGCCTCACTCTATCTCATTATGTGTTTACGATGAATTAGTGGATTCTTGC 1281
Qy        401 LysProGlyAspArgIleGluValThrGlyValPheLysAlaMetAlaValArgValGly 420
Db       1282 AGGGCGGGCGATCGTATTGAGGTGACTGGCACGTTCAGGTCCATCCCCATTAGAGCTAAT 1341
Qy        421 ProAsnGlnArgThrLeuArgAlaLeuTyrLysThrTyrIleAspCysValHisValLys 440
Db       1342 TCCAGGCAACGCGTACTAAAGTCGTTGTATAAAACATACGTCGATGTGGTCCACGTTAAA 1401
Qy        441 LysSerAspArgGlyArgLeuGlnThrGluAspProMetGluMetAspLysGluAsnAsp 460
Db       1402 AAAGTTTCAGATAAGAGGTTAGACGTCGATACTTCTACTATTGAACAAGAATTAATGCAG 1461
Qy        461 MetTyrAlaGlyTyrHisGluSerAspThrSerGluAlaAlaAsnGluAlaLysIleGln 480
Db       1462 AACAAGGTAGATCATAACGAGGTCGAAGAAGTAAGACAAATTACTGATCAGGATTTAGCA 1521
Qy        481 LysLeuLysGluLeuSerLysLeuProGlyIleTyrAspArgLeuSerArgSerLeuAla 500
Db       1522 AAAATTCGCGAGGTTGCGGCAAGAGAAGATTTATACAGTTTATTAGCCCGCTCTATTGCC 1581
Qy        501 ProSerIleTrpGluLeuGluAspIleLysLysGlyLeuLeuCysGlnLeuPheGlyGly 520
Db       1582 CCAAGTATTTACGAGCTAGAAGATGTCAAGAAGGGTATATTACTTCAGCTATTTGGCGGC 1641
Qy        521 LysAlaLysLysIleProSerGlyAlaSerPheArgGlyAspIleAsnValLeuLeuVal 540
Db       1642 ACGAATAAAACCTTTACCAAGGGTGGCCGTTATAGAGGTGACATAAATATTTTACTTTGT 1701
Qy        541 GlyAspProGlyThrSerLysSerGlnLeuLeuGlnTyrValHisLysIleAlaProArg 560
Db       1702 GGGGATCCTTCTACTTCCAAATCGCAAATTTTGCAATACGTTCACAAAATTACTCCTCGT 1761
Qy        561 GlyIleTyrThrSerGlyArgGlySerSerAlaValGlyLeuThrAlaTyrValThrLys 580
Db       1762 GGTGTGTATACTTCGGGTAAAGGTTCATCTGCCGTTGGTTTAACTGCTTATATTACAAGG 1821
Qy        581 AspProGluThrArgGluThrValLeuGluSerGlyAlaLeuValLeuSerAspArgGly 600
Db       1822 GATGTCGACACAAAACAACTTGTTTTGGAAAGTGGTGCATTAGTATTGTCTGATGGAGGT 1881

Qy        601 IleCysCysIleAspGluPheAspLysMetSerAspAsnAlaArgSerMetLeuHisGlu 620
Db       1882 GTTTGTTGTATTGACGAGTTTGATAAAATGAGTGATTCTACAAGGTCCGTCTTGCACGAA 1941
Qy        621 ValMetGluGlnGlnThrValSerValAlaLysGlyGlyIleIleAlaSerLeuAsnAla 640
Db       1942 GTCATGGAACAGCAGACTATTTCAATCGCAAAAGCGGGAATTATCACAACACTTAATGCC 2001
Qy        641 ArgThrSerValLeuAlaCysAlaAsnProSerGlySerArgTyrAsnAlaArgLeuSer 660
Db       2002 AGAAGTTCTATTTGGCCAGTGCTAACCCAATTGGTTCACGCTACAACCCCAATTTGCCT 2061
Qy        661 ValIleAspAsnIleGlnLeuProProThrLeuLeuSerArgPheAspLeuIleTyrLeu 680
Db       2062 GTGACTGAAAATATTGATCTACCGCCCCCACTACTTTCGAGATTCGATCTGGTCTATCTT 2121
Qy        681 MetLeuAspLysProAspGluGlnAsnAspArgArgLeuAlaArgHisLeuValAlaLeu 700
Db       2122 GTCCTTGATAAGGTTGATGAGAAAAATGACAGAGAACTAGCCAAACACTTAACAAATCTT 2181
Qy        701 HisTyrGluAsn------TyrGluValSerLysGlnAspAlaLeuAspLeuGlnThrLeu 718
Db       2182 TACCTGGAAGATAAGCCCGAACATATTTCTCAAGACGACGTGCTACCAGTTGAATTTTTA 2241
Qy        719 ThrAlaTyrIleThrTyrAlaArgGlnHisValHisProThrLeuSerAspGluAlaAla 738
Db       2242 ACGATGTATATTAGTTATGCAAAGGAGCACATACACCCAATAATCACCGAGGCCGCTAAG 2301
Qy        739 GluAspLeuIleAsnGlyTyrValGluMetArgGlnLysGlyAsnPheProGlySerSer 758
Db       2302 ACTGAGCTTGTTCGTGCTTATGTAGGAATGAGAAAGATGGGTGACGATTCGAGATCCGAT 2361
Qy        759 LysLysValIleThrAlaThrProArgGlnLeuGluSerMetIleArgIleSerGluAla 778
Db       2362 GAGAAGAGAATCACAGCTACCACAAGACAACTTGAAAGTATGATTCGTTTGGCTGAGGCG 2421
Qy        779 LeuAlaArgMetArgPheSerGluValValGluLysValAspAlaAlaGluAlaValArg 798
Db       2422 CACGCCAAAATGAAATTGAAAAACGTCGTAGAGCTGGAGGATGTTCAAGAAGCCGTTAGA 2481
Qy        799 LeuLeuAspValAlaLeuGlnGlnSerAlaThrAspHisAlaThrGlyThrIleAspMet 818
Db       2482 TTAATTAGATCAGCCATAAAAGATTATGCAACAGACCCTAAAACCGGTAAAATCGACATG 2541
Qy        819 AspLeuIleThrThrGlyValSerAlaSerGluArgIleArgArgAlaAsnLeuLeuAla 838
Db       2542 AATTTAGTTCAAACAGGTAAATCAGTTATTCAGAGAAAACTACAGGAGGATTTGTCAAGG 2601
Qy        839 AlaLeuArgGluLeuIleAlaAspLysIleSerProGlySerSerSerGlyLeuLysThr 858
Db       2602 GAAATTATGAATGTTTTGAAGGATCAG------------GCATCAGACTCAATGTCATTC 2649
Qy        859 SerGlnLeuLeuGluAspIleArgSerGlnSerSerValAspValSerLeuGlnAspIle 878
Db       2650 AATGAGCTTATAAAACAAATCAATGAACACTCTCAAGATAGAGTTGAGTCTTCTGATATC 2709
Qy        879 LysAsnAlaLeuGlySerLeuGlnGlyGluGlyPheLeuThrValHisGlyAspIleVal 898
Db       2710 CAGGAAGCCTTGTCAAGATTGCAACAAGAGGACAAGGTCATTGTCCTTGGCGAGGGTGTA 2769
Qy        899 LysArg 900
Db       2770 AGGAGA 2775
```

RESULT 9
ADS49058
ID ADS49058 standard; cDNA; 2881 BP.
XX
AC ADS49058;
XX
DT 02-DEC-2004 (first entry)
XX
DE Bacterial polynucleotide #3801.

```
XX
KW   Recombinant DNA construct; transformed plant; improved plant property;
KW   cold tolerance; heat tolerance; drought tolerance; herbicide; osmosis;
KW   pathogen tolerance; pest tolerance; plant disease resistance;
KW   cell cycle pathway modification; plant growth regulator;
KW   homologous recombination; seed oil yield; protein yield; carbohydrate;
KW   nitrogen; phosphorus; photosynthesis; lignin; galactomannan;
KW   bacterial polynucleotide; gene; ss.
XX
OS   Bacteria.
XX
PN   US2003233675-A1.
XX
PD   18-DEC-2003.
XX
PF   20-FEB-2003; 2003US-00369493.
XX
PR   21-FEB-2002; 2002US-0360039P.
XX
PA   (CAOY/) CAO Y.
PA   (HINK/) HINKLE G J.
PA   (SLAT/) SLATER S C.
PA   (CHEN/) CHEN X.
PA   (GOLD/) GOLDMAN B S.
XX
PI   Cao Y,  Hinkle GJ,  Slater SC,  Chen X,  Goldman BS;
XX
DR   WPI; 2004-061375/06.
XX
PT   New recombinant DNA construct comprising a promoter positioned to provide
PT   for expression of a polynucleotide encoding a polypeptide from a
PT   microbial source, useful for producing plants with improved properties.
XX
PS   Claim 1; SEQ ID NO 27488; 122pp; English.
XX
CC   The invention relates to a recombinant DNA construct comprising a
CC   promoter functional in a plant cell, where the promoter is positioned to
CC   provide for expression of a polynucleotide encoding a polypeptide from a
CC   microbial source. The invention also relates to a transformed plant
CC   comprising the recombinant DNA construct and a method of producing a
CC   transformed plant having an improved property. The plant is a crop plant
CC   such as maize or soybean. The method of producing a transformed plant
CC   having an improved property comprises transforming a plant with the
CC   recombinant DNA construct and growing the transformed plant, where the
CC   polynucleotide or polypeptide is useful for improving plant properties.
CC   The recombinant DNA construct is useful for producing plants with
CC   improved plant properties, e.g. improved cold, heat or drought tolerance,
CC   tolerance to herbicides, extreme osmotic conditions, pathogens or pests,
CC   increased resistance to plant disease, better growth rate by modification
CC   of the cell cycle pathway with plant growth regulators, increased rate of
CC   homologous recombination, modified seed oil or protein yield and/or
CC   content, improved yield by modification of carbohydrate, nitrogen or
CC   phosphorus use and/or uptake, by modification of photosynthesis or by
CC   providing improved plant growth and development under at least one stress
CC   condition, improved lignin production or improved galactomannan
CC   production. This sequence represents a bacterial polynucleotide used in
CC   the scope of the invention. Note: The sequence data for this patent did
CC   not form part of the printed specification but was obtained in electronic
CC   format from USPTO at seqdata.uspto.gov/sequence.html.
XX
SQ   Sequence 2881 BP; 678 A; 858 C; 746 G; 599 T; 0 U; 0 Other;

Alignment Scores:
Pred. No.:              4.5e-98         Length:         2881
Score:                  1819.00         Matches:        419
Percent Similarity:     58.17%          Conservative:   161
Best Local Similarity:  42.03%          Mismatches:     249
Query Match:            39.47%          Indels:         169
DB:                     13              Gaps:           24

US-10-768-511-8 (1-901) x ADS49058 (1-2881)

Qy         24 SerThrProLeuProGlnValThrSerProSerPheAsp-------------------- 36
                 |||:::||||||   |||    |||                |||
Db          1 AGCTCGCCTTTGCGCCAGCAGACAGAGACCCAGAGCGATGCTGACCGAACTCCTCGCGCA 60

Qy         37 AsnAlaAlaSerProValAlaGlyArgArgAlaValArgGlnThr--------------- 51
              |||     |||  :::  |||     ||||||  |||
Db         61 AACGGACGGAGCCAGTTAATTGGGGGTATGGCTGTCAACCAAAAAAAAAAAAAAAAGGAC 120

Qy         52 ---------------ProThrSerAlaValArgArgArgGlyArgGluThrAspSerAla 66
                             |||  ||||||                       ||||||:::
Db        121 TACACTGAATGCTGACCACACTCCGCA-----------------------GACTCCTCT 156

Qy         67 ArgArgArgArgSerArgSerArgSerLeuGlyAsn-SerValTyrSerSerProTyr-- 85
                 |||  :::  |||  |||  |||  |||:::   :::||||||
Db        157 CCCATCCGCTATGCCTCCAGTTCTAGCCCGGGCCGTCAGCTTACTCAACAGTCCGATCTT 216

Qy         86 -------------AspAlaGlyThrProGlyThrProGlyThrProValAlaThrProVa 101
                          :::  :::              |||                |||||
Db        217 CGTAGCGAAAGCAGTCAGCTCTTCGTCAGCTCGCAAAGGTCCGTTGCCGGTCGATCCCGT 276

Qy        101 lTyrAlaThrProValGlyThrProMetGlyThrProSerPheHisArgGlyThrProGl 121
              |  ||||||     |||   |||:::  ||||||:::                    ||
Db        277 CGAGGCGACATCAACGGGGAC-CCTCTTCGTACTCCCGCC------------------CA 317

Qy        121 nTyrLysGlnArgSerGluLeuGlySerGlnGlyLysProLeuHisArgArgArgArgSe 141
              |     :::|||     |||        |||:::  :::
Db        318 GATTCCCCGCCGAATCATCCTCGACGACGCCGGTCGTGTGATCCGC-------------- 363

Qy        141 rGlnSerArgGluProGlyHisArgSerProSerArgGluProSerAlaAspGlyArgPr 161
                                                        |||   :::|||
Db        364 ---------------------------------GATGCCCCTGGCTCCGATGCCAACTC 389

Qy        161 oSerGluSerAlaGluProAsp------------AspThrLeuGlyGlyGlu---TyrAl 176
                 :::          :::           |||:::   ||||||||||:::
Db        390 GTTTGTCACCAATAACCCCAATACCTCTGAGGCAGATGCCCTTGGTGGACAAAGCCAAGG 449

Qy        176 aTyrValTrpGlyThrAsnValAsnIleProAspValLeuArgAlaIleArgArgPheLe 196
                 ||||||||||||   :::::::::  |||         |||   :::   |||||
Db        450 TCTTGTTTGGGGTACAACAATTTCTCTTGACGATTCGTTCTCCGCGTTCAAGGACTTTCT 509

Qy        196 uHisAsn-----------TyrArgSerSerAlaHis----------------------- 204
              |   |||           ||||||      |||
Db        510 CAGGAACTTCACCAGGAAATACCGGATGTGGGCAGATGGCGCGGACGAGGCCGAGACTAT 569

Qy        205 ----------AspLeuAsnSerLys---TyrIleGlnIleIleGluGluThrValGluAr 220
                         |||  :::||||||   |||   :::   :::|||     :::
Db        570 TGGCCATCCGGATGCGGATTCGAAGCCATATTGGGAGGCACTGGAGAATATGCTGCTGCT 629

Qy        220 gGluGluAspThrLeuAsnIleAspMetSerAspIle------TyrAspHisAspProAs 238
                 :::   |||   :::|||:::   |||:::         |||
Db        630 TGGTACCAACAAGCTGTATCTGGATCTCCGCGACCTCAAGTCCTACCCGCGCACCTTGAA 689

Qy        238 pLeuTyrAlaLysIleValArgTyrProLeuAspIleIleProLeuLeuAspThr----- 256
                |||:::   :::         ||||||   :::|||||||||||::::::|||
Db        690 GTTATGGCATCAGGCCCAACACTATCCTACCGAAATCATCCCGGTCATGGATCAATGCGT 749

Qy        257 ----GluCys-----------------GlnGluValAlaThr----------------- 263
                  :::|||                 :::|||:::|||:::
Db        750 ACATGACTGTATGATGGAGTTGGCGCAAAAGGAAATGGCCAGCCAAAGAGCATCCCAAAA 809

Qy        263 ------------------------------------------------------------ 263

Db        810 CTCTAGGACCGCTCCAGGTGCCTCACAAAGCTCGGAGCCGAACTTCCCAAGCTCTGAAAG 869

Qy        264 ---------------------------------SerLeuLeuProThrPheGluLysHi 272
                                               :::      ||||||   |||
Db        870 GAGCGAGGAACCGCCCACCCCGCGGCCTGCCCAAACAGCGGCGCCCACAATTGAGGACCA 929

Qy        272 sIleGlu---------------AlaArgProPheAsnLeuLysAlaSerValHisMetAr 287
               :::                  ||||||:::   |||              ::::::||
Db        930 AGTCAGCCAGATGGCGTATGTGGTTCGCCCGTGGGGTCTTGACAAGATCACCAACCTCCG 989

Qy        287 gGluLeuAsnProSerAspIleAspLysLeuValSerValLysGlyMetValIleArgCy 307
              |:::||||||||||||||||:::||||||||||||||||:::||||||:::|||||||||
```

```
Db    990 TGACCTCAACCCATCTGACATGGACAAGCTTGTGTCCATCAAGGGTCTTGTCATCAGAAC 1049
Qy    307 sSerSerIleIleProGluIleLysGlyAlaPhePheLysCysLeuValCysGlyHisSe 327
Db   1050 CACGCCGGTCATCCCCGACATGAAGGATGCATTTTTCAAGTGCAGTGTTTGTGGTCATTC 1109
Qy    327 rProProLeuValThrValVal------LysGlyArgValGluGluProThrArgCysGl 345
Db   1110 T---------ATAACCGTTCAGTTAGATCGTGGCAAGATTCGTGAGCCTACAGAGTGTCC 1160
Qy    345 uLysProGluCysAlaAlaArgAsnAlaMetSerLeuIleHisAsnArgCysThrPheAl 365
Db   1161 TCGGGCTCGCTGCGCCTCCAAGAACTCAATGCAGATCATCCACAACAGATGTGCTTTCGA 1220
Qy    365 aAsnLysGlnIleValArgLeuGlnGluThrProAspAlaIleProGluGlyGluThrPr 385
Db   1221 AGACAAGCAAGTCATCAAGCTGCAAGAAACTCCGGATAACGTGCCTGCTGGTCAGACACC 1280
Qy    385 oHisThrValSerMetCysLeuTyrAsnThrMetValAspAlaValLysProGlyAspAr 405
Db   1281 CCACTCGGTATCCGTTTGCGTCTATAACGAGCTTGTGGACTTTTGCAAGGCCGGTGACAG 1340
Qy    405 gIleGluValThrGlyValPheLysAlaMetAlaValArgValGlyProAsnGlnArgTh 425
Db   1341 AGTTGAGCTCACTGGTATCTTTAAAGTCACCCCCGTTCGTGTCAACCCGCGCATGAGGAC 1400
Qy    425 rLeuArgAlaLeuTyrLysThrTyrIleAspCysValHisValLysLysSerAspArgGl 445
Db   1401 GGTCAAGAGCGTGCACAAGACATATGTTGATGTTGTCCATGTACAAAAGGTCGACCGCAA 1460
Qy    445 yArgLeuGlnThrGluAspProMetGluMetAsp------LysGluAsnAspMetTyrAl 463
Db   1461 GAGGATGGGTTCT---GATCCCTCGACCCTCGACCTTGCCGAGGAGGAGGAAGCCCATGC 1517
Qy    463 aGlyTyrHisGluSerAspThrSerGluAlaAlaAsnGluAlaLysIleGlnLysLeuLy 483
Db   1518 TAATGGCCAGAGCATGGACGAAGTCCGAAAGGTCTCTCCTGATGAGGAAGAGAGGATCAA 1577
Qy    483 sGluLeuSerLysLeuProGlyIleTyrAspArgLeuSerArgSerLeuAlaProSerIl 503
Db   1578 GGAAACTGCCGCTCGACCGGATATCTACGATCTCCTTTCACGTTCTTTGGCTCCATCAAT 1637
Qy    503 eTrpGluLeuGluAspIleLysLysGlyLeuLeuCysGlnLeuPheGlyGlyLysAlaLy 523
Db   1638 TTACGAGATGGATGATGTGAAGAAGGGTATTCTTCTTCAACTGTTCGGAGGCACGAACAA 1697
Qy    523 sLysIleProSerGlyAlaSer------PheArgGlyAspIleAsnValLeuLeuValGl 541
Db   1698 GACCTTTGAGAAGGGTGGAAGCCCCAAATACCGTGGAGACATCAACGTTCTCCTCTGTGG 1757
Qy    541 yAspProGlyThrSerLysSerGlnLeuLeuGlnTyrValHisLysIleAlaProArgGl 561
Db   1758 TGACCCATCCACTTCCAAGTCCCAGCTTCTCTCCTATGTTCACAGAATCGCCCCAAGAGG 1817
Qy    561 yIleTyrThrSerGlyArgGlySerSerAlaValGlyLeuThrAlaTyrValThrLysAs 581
Db   1818 TGTGTACACGAGTGGTAAAGGTTCTTCTGCCGTAGGTCTTACCGCCTATGTTACCCGCGA 1877
Qy    581 pProGluThrArgGluThrValLeuGluSerGlyAlaLeuValLeuSerAspArgGlyIl 601
Db   1878 TCCCGAGTCCCGCCAACTCGTTCTAGAGTCCGGTGCTCTCGTCTTATCCGACGGCGGTGT 1937
Qy    601 eCysCysIleAspGluPheAspLysMetSerAspAsnAlaArgSerMetLeuHisGluVa 621
Db   1938 CTGTTGTATCGACGAGTTCGACAAGATGAACGAATCGACCCGCTCCGTTCTTCACGAAGT 1997
Qy    621 lMetGluGlnGlnThrValSerValAlaLysGlyGlyIleIleAlaSerLeuAsnAlaAr 641
Db   1998 CATGGAACAACAGACCGTGTCCGTCGCCAAGGCCGGCATCATCACCACTCTCAATGCCCG 2057
Qy    641 gThrSerValLeuAlaCysAlaAsnProSerGlySerArgTyrAsnAlaArgLeuSerVa 661
Db   2058 AACATCAATCCTGGCGTCTGCCAACCCCATTGGTAGTAGATACAACCCTGATCTATCAGT 2117
```

```
Qy    661 lIleAspAsnIleGlnLeuProProThrLeuLeuSerArgPheAspLeuIleTyrLeuMe 681
Db   2118 GCCGCAAAATATTGACCTCCCGCCAACCCTTCTATCCCGTTTCGATCTGGTCTACCTCAT 2177
Qy    681 tLeuAspLysProAspGluGlnAsnAspArgArgLeuAlaArgHisLeuValAlaLeuHi 701
Db   2178 CCTCGATCGTGTTGATGAGAAGAATGATCAGCGTCTGGCCCGCCATTTGCTCTCCATGTA 2237
Qy    701 sTyrGluAsnTyrGluValSerLysGlnAspAlaLeuAspLeu---------------- 715
Db   2238 TCTCGAAGACAAGCCCGAGAGTGCCCAGCAGGCCAACGACGTCCTTGTGAGTTTTTCCCC 2297
Qy    715 ------------------------------------------------------------ 715
Db   2298 GTCTGTCAATCCACACGAAGATGAACCCCAAACTAACCAACTCCCCCTCAAAAAAACAGC 2357
Qy    716 ----GlnThr-LeuThrAlaTyrIleThrTyrAlaArgGlnHisValHisProThrLeuS 734
Db   2358 CCGTCGAGTTTCTAACCTCCTACATCTCCTACGCCCGCTCGCACATCCACCCGGCCCTCA 2417
Qy    734 erAspGluAlaAlaGluAspLeuIleAsnGlyTyrValGluMetArgGlnLysGlyAsnP 754
Db   2418 CGCCCGAAGCCGGTCGCGAGCTCGTCGATGCCTACGTCGAGATGCGCAAGCTGGGCCAAG 2477
Qy    754 heProGlySerSerLysLysValIleThrAlaThrProArgGlnLeuGluSerMetIleA 774
Db   2478 ACGTGCGCGCCGCCGAGAAGCGCATTACGGCCACCACACGTCAGCTCGAGTCCATGATTC 2537
Qy    774 rgIleSerGluAlaLeuAlaArgMetArgPheSerGluValValGluLysValAspAlaA 794
Db   2538 GCTTAGCCGAGGCGCACGCAAAGATGCGTTTATCTCAAACCGTCACGCGCGACGACGTCC 2597
Qy    794 laGluAlaValArgLeuLeuAspValAlaLeuGlnGlnSerAlaThrAspHisAlaThrG 814
Db   2598 GCGAGGCCGTGCGCCTCATCAAGTCGGCGCTCAAGACCGCCGCCACCGAC---AGCCAGG 2654
Qy    814 lyThrIleAspMetAspLeuIleThrThrGlyValSerAlaSerGluArgIleArgArgA 834
Db   2655 GCCGCATCGACATGAGCTTGTTGACAGAGGGCACGAGCGCGGCGGAGCGCCAGCGCAAGG 2714
Qy    834 laAsnLeuLeuAlaAlaLeuArgGluLeuIleAlaAspLysIleSerProGlySerSerS 854
Db   2715 CCGACATGAAGGATGCGGTTATCCGCCTGCTGGATGAGATGACCAGCGGCGGACAGGTA- 2773
Qy    854 erGlyLeuLysThrSerGlnLeuLeuGluAspIleArgSerGlnSerSerValAspValS 874
Db   2774 -----GTTAGGTACTCGGAGGTGGCGAGACGGTTGGGCGAGGGGGCGGGTGTGCAGGTTG 2828
Qy    874 erLeuGlnAspIleLysAsnAlaLeuGlySerLeuGlnGlyGluGly 889
Db   2829 AGCCGGCTGAGTTTGCGGAGGTTATGAGGGCTCTGGAGATGGAGGGC 2875
```

RESULT 10
ABL65258
ID   ABL65258 standard; DNA; 3273 BP.
XX
AC   ABL65258;
XX
DT   15-MAY-2002 (first entry)
XX
DE   Lung cancer related gene sequence SEQ ID NO:3595.
XX
KW   Human; cancer; colon; breast; ovary; oesophagus; kidney; thyroid;
KW   stomach; lung; prostate; pancreas; carcinoma; antitumour; cancerous;
KW   cytostatic; gene therapy; antineoplastic; Wilm's tumour; adenocarcinoma;
KW   gene; ds.
XX
OS   Homo sapiens.
XX
PN   WO200194629-A2.
XX
PD   13-DEC-2001.

```
XX
PF   30-MAY-2001; 2001WO-US010838.
XX
PR   05-JUN-2000; 2000US-0209473P.
PR   05-JUN-2000; 2000US-0209531P.
PR   18-SEP-2000; 2000US-0233133P.
PR   18-SEP-2000; 2000US-0233617P.
PR   20-SEP-2000; 2000US-0234009P.
PR   20-SEP-2000; 2000US-0234034P.
PR   20-SEP-2000; 2000US-0234052P.
PR   22-SEP-2000; 2000US-0234509P.
PR   22-SEP-2000; 2000US-0234567P.
PR   25-SEP-2000; 2000US-0234923P.
PR   25-SEP-2000; 2000US-0234924P.
PR   25-SEP-2000; 2000US-0235077P.
PR   25-SEP-2000; 2000US-0235082P.
PR   25-SEP-2000; 2000US-0235134P.
PR   25-SEP-2000; 2000US-0235280P.
PR   26-SEP-2000; 2000US-0235637P.
PR   26-SEP-2000; 2000US-0235638P.
PR   27-SEP-2000; 2000US-0235711P.
PR   27-SEP-2000; 2000US-0235720P.
PR   27-SEP-2000; 2000US-0235840P.
PR   27-SEP-2000; 2000US-0235863P.
PR   28-SEP-2000; 2000US-0236028P.
PR   28-SEP-2000; 2000US-0236032P.
PR   28-SEP-2000; 2000US-0236033P.
PR   28-SEP-2000; 2000US-0236034P.
PR   28-SEP-2000; 2000US-0236109P.
PR   28-SEP-2000; 2000US-0236111P.
PR   29-SEP-2000; 2000US-0236842P.
PR   29-SEP-2000; 2000US-0236891P.
PR   02-OCT-2000; 2000US-0237172P.
PR   02-OCT-2000; 2000US-0237173P.
PR   02-OCT-2000; 2000US-0237278P.
PR   02-OCT-2000; 2000US-0237294P.
PR   02-OCT-2000; 2000US-0237295P.
PR   02-OCT-2000; 2000US-0237316P.
PR   03-OCT-2000; 2000US-0237425P.
PR   03-OCT-2000; 2000US-0237598P.
PR   03-OCT-2000; 2000US-0237604P.
PR   03-OCT-2000; 2000US-0237606P.
PR   03-OCT-2000; 2000US-0237608P.
PR   01-NOV-2000; 2000US-0244867P.
PR   01-NOV-2000; 2000US-0245084P.
XX
PA   (AVAL-) AVALON PHARM.
XX
PI   Young PE,  Augustus M,  Carter KC,  Ebner R,  Endress G,  Horrigan S;
PI   Soppet DR,  Weaver Z;
XX
DR   WPI; 2002-188264/24.
XX
PT   Screening for anti-neoplastic agent involves exposing cells to a chemical
PT   agent to be tested for anti-neoplastic activity, and determining a change
PT   in expression of a gene of a signature gene set.
XX
PS   Claim 1; SEQ ID NO 3595; 44pp; English.
XX
CC   The present invention describes a method (M1) for screening for an anti-
CC   neoplastic agent. The method involves exposing cells to a chemical agent
CC   to be tested for anti-neoplastic activity, determining a change in
CC   expression of at least one gene (I) of a signature gene set, where (I)
CC   comprises a sequence (S) selected from 8447 sequences (given in ABL61664
CC   to ABL70110), or is at least 95% identical to (S), where a change in
CC   expression is indicative of anti-neoplastic activity. (I) has cytostatic
CC   activity and can be used in gene therapy. M1 can be used for screening an
CC   anti-neoplastic agent, and can be used for producing a product which is
CC   the data collected with respect to the anti-neoplastic agent as a result
CC   of M1, and the data is sufficient to convey the chemical structure and/or
CC   properties of the agent. M1 can be used in the treatment of cancer such
CC   as colon, breast, stomach, lung, thyroid, oesophageal, ovarian, kidney,
CC   prostate or pancreatic cancer, adenocarcinoma, carcinoma, clear cell
CC   cancer, infiltrating ductal cancer, infiltrating lobular cancer, squamous
CC   cell carcinoma, neuroendocrine carcinoma, papillary carcinoma and Wilm's
CC   tumour
XX
SQ   Sequence 3273 BP; 876 A; 809 C; 859 G; 729 T; 0 U; 0 Other;

Alignment Scores:
Pred. No.:              6.4e-98        Length:        3273
Score:                  1817.50        Matches:       406
Percent Similarity:     57.78%         Conservative:  151
Best Local Similarity:  42.12%         Mismatches:    256
Query Match:            39.43%         Indels:        151
DB:                     6              Gaps:          18

US-10-768-511-8 (1-901) x ABL65258 (1-3273)

Qy         38 AlaAlaSerProValAlaGlyArgArgAlaValArgGlnThrProThrSerAlaValArg 57
              |||   |||       :::||||||                   |||   :::|||   |||
Db         21 GCGACGTCGTCGGCAAGCGGCCGC--------------CTTCCACGTAACGCG---CGC 62

Qy         58 ArgArgGlyArgGluThrAspSerAlaArgArgArgArgSerArg--------------- 72
              |||||||||||             :::||||||  |||  |||
Db         63 CGGCGGGGGAGGGCGTTGGCGCGGAGCCGACGGGAACGTCCGCGCTGCGGAGCAGGGCAG 122

Qy         73 ---SerArgSerLeuGlyAsnSerValTyrSerSerProTyrAspAlaGlyThrProGly 91
                 ||||||      |||                :::   |||      |||||||||||
Db        123 GGAAGCCGGGAGGCGGGC---CCGGCCCGAGCTTGTCCTTGTCGCGCAGGTACTCCGAGC 179

Qy         92 ThrProGlyThrProValAlaThrProVal--------------------TyrAlaThr 104
              |||     :::|||   :::||||||                          ||||||
Db        180 ACTATGTCGTCCCCGGCGTCGACCCCGAGCCGCCGCGGCAGCCGGCGTGGAAGGGCCACC 239

Qy        105 ProValGlyThrPro--------------------------------------------- 109
              |||     ||||||
Db        240 CCCGCCCAGACGCCTCGGAGTGAGGATGCCAGGTCATCTCCCTCTCAGAGACGTAGAGGC 299

Qy        109 ------------------------------------------------------------ 109

Db        300 GAGGATTCCACCTCCACGGGGGAGTTGCAGCCGATGCCAACCTCGCCTGGAGTGGACCTG 359

Qy        109 ------------------------------------------------------------ 109

Db        360 CAGAGCACTGCTGCGCAGGACGTGCTGTTTTCCAGCCCTCCCCAAATGCATTCTTCAGCT 419

Qy        110 ------------------------------------MetGlyThrProSerPheHis--- 116
                                                   ||||||||||||
Db        420 ATCCCTCTTGACTTTGATGTTAGTTCACCACTGACATACGGCACTCCCAGCTCTCGGGTA 479

Qy        117 ------------------------ArgGlyThrProGlnTyrLysGlnArgSerGluLeu 128
                                      ||||||||||||      :::||||||   :::|||
Db        480 GAGGGAACCCCAAGAAGTGGTGTTAGGGGCACACCT---GTGAGACAGAGGCCTGACCTG 536

Qy        129 GlySerGlnGlyLysProLeuHisArgArgArgArgSerGlnSerArgGluProGlyHis 148
              ||||||      |||   |||               :::|||
Db        537 GGCTCTGCACAGAAGGGCCTGCAAGTGGATCTGCAGTCT--------------------- 575

Qy        149 ArgSerProSerArgGluProSerAlaAspGlyArgProSerGluSerAlaGluProAsp 168
                                          ||||||      :::|||
Db        576 ---------------------------GACGGGGCAGCAGCAGAAGATATAGTGGCAAGT 608

Qy        169 AspThrLeuGlyGlyGluTyrAlaTyrValTrpGlyThrAsnValAsnIleProAspVal 188
              :::            |||:::          :::|||||||||:::||||||:::
Db        609 GAGCAGTCTCTAGGCCAAAAACTTGTGATCTGGGGAACAGATGTAAATGTGGCAGCATGC 668

Qy        189 LeuArgAlaIleArgArgPheLeuHisAsnTyr-----------------------Arg 200
                          :::|||||||||      :::
Db        669 AAAGAAAACTTTCAGAGATTTCTTCAGCGTTTTATTGACCCTCTGGCTAAAGAAGAAGAA 728

Qy        201 SerSerAlaHisAspLeuAsnSerLys---TyrIleGlnIleIleGluGluThrValGlu 219
              :::            |||:::             |||:::|||   :::   |||
Db        729 AATGTTGGCATAGATATTACTGAACCTCTATACATGCAACGACTTGGGGAGATTAATGTT 788

Qy        220 ArgGluGluAspThrLeuAsnIleAspMetSerAspIleTyrAspHisAspProAspLeu 239
```

```
Db        789 ATTGGTGAGCAATTTTTAAATGTGAACTGTGAACACATCAAATCATTTGACAAAAATTTG 848

Qy        240 TyrAlaLysIleValArgTyrProLeuAspIleIleProLeuLeuAspThrGluCysGln 259
Db        849 TACAGACAACTCATCTCTTACCCACAGGAAGTTATTCCAACTTTTGACATGGCTGTCAAT 908

Qy        260 GluValAlaThrSerLeuLeuPro------ThrPheGluLysHisIleGluAlaArgPro 277
Db        909 GAAATCTTCTTTGACCGTTACCCTGACTCAATCTTAGAACATCAGATTCAAGTAAGACCA 968

Qy        278 PheAsnLeuLysAlaSerValHisMetArgGluLeuAsnProSerAspIleAspLysLeu 297
Db        969 TTCAACGCATTGAAGACTAAGAATATGAGAAACCTGAATCCAGAAGACATTGACCAGCTC 1028

Qy        298 ValSerValLysGlyMetValIleArgCysSerSerIleIleProGluIleLysGlyAla 317
Db       1029 ATCACCATCAGCGGCATGGTGATCAGGACATCCCAGCTGATTCCCGAGATGCAGGAGGCC 1088

Qy        318 PhePheLysCysLeuValCysGlyHisSerProProLeuValThrValValLysGlyArg 337
Db       1089 TTCTTCCAGTGCCAAGTGTGTGCCCACACG---ACCCGGGTGGAGATGGACCGCGGCCGC 1145

Qy        338 ValGluGluProThrArgCysGluLysProGluCysAlaAlaArgAsnAlaMetSerLeu 357
Db       1146 ATTGCAGAGCCCAGTGTGTGCGGGCGC------TGCCACACCACCCACAGCATGGCACTC 1199

Qy        358 IleHisAsnArgCysThrPheAlaAsnLysGlnIleValArgLeuGlnGluThrProAsp 377
Db       1200 ATCCACAACCGCTCCCTCTTCTCTGACAAGCAGATGATCAAGCTTCAGGAGTCTCCGGAA 1259

Qy        378 AlaIleProGluGlyGluThrProHisThrValSerMetCysLeuTyrAsnThrMetVal 397
Db       1260 GACATGCCTGCAGGGCAGACACCACACACAGTTATCCTGTTTGCTCACAATGATCTCGTT 1319

Qy        398 AspAlaValLysProGlyAspArgIleGluValThrGlyValPheLysAlaMetAlaVal 417
Db       1320 GACAAGGTCCAGCCTGGGGACAGAGTGAATGTTACAGGCATCTATCGAGCTGTGCCTATT 1379

Qy        418 ArgValGlyProAsnGlnArgThrLeuArgAlaLeuTyrLysThrTyrIleAspCysVal 437
Db       1380 CGAGTCAATCCAAGAGTGAGTAATGTGAAGTCTGTCTACAAAACCCACATTGATGTCATT 1439

Qy        438 HisValLysLysSerAspArgGlyArgLeuGlnThrGluAspProMetGluMetAspLys 457
Db       1440 CATTATCGGAAAACGGATGCAAAACGTCTGCAT-------------------------- 1472

Qy        458 GluAsnAspMetTyrAlaGlyTyrHisGluSerAspThrSerGluAlaAlaAsnGluAla 477
Db       1473 -----------------GGCCTTGATGAAGAAGCAGAACAGAAACTTTTTTCAGAGAAA 1514

Qy        478 LysIleGlnLysLeuLysGluLeuSerLysLeuProGlyIleTyrAspArgLeuSerArg 497
Db       1515 CGTGTGGAATTGCTTAAGGAACTTTCCAGGAAACCAGACATTTATGAGAGGCTTGCTTCA 1574

Qy        498 SerLeuAlaProSerIleTrpGluLeuGluAspIleLysLysGlyLeuLeuCysGlnLeu 517
Db       1575 GCCTTGGCTCCAAGCATTTATGAACATGAAGATATAAAGAAGGGAATTTTGCTTCAGCTC 1634

Qy        518 PheGlyGlyLysAlaLysLysIle------ProSerGlyAlaSerPheArgGlyAspIle 535
Db       1635 TTTGGCGGGACAAGGAAGGATTTTAGTCACACTGGAAGGGGCAAATTTCGGGCTGAGATC 1694

Qy        536 AsnValLeuLeuValGlyAspProGlyThrSerLysSerGlnLeuLeuGlnTyrValHis 555
Db       1695 AACATCTTGCTGTGTGGCGACCCTGGTACCAGCAAGTCCCAGCTGCTGCAGTACGTGTAC 1754

Qy        556 LysIleAlaProArgGlyIleTyrThrSerGlyArgGlySerSerAlaValGlyLeuThr 575
Db       1755 AACCTCGTCCCCAGGGGCCAGTACACGTCTGGGAAGGGCTCCAGTGCAGTTGGCCTCACT 1814

Qy        576 AlaTyrValThrLysAspProGluThrArgGluThrValLeuGluSerGlyAlaLeuVal 595
Db       1815 GCGTACGTAATGAAAGACCCTGAGACAAGGCAGCTGGTCCTGCAGACAGGTGCTCTTGTC 1874

Qy        596 LeuSerAspArgGlyIleCysCysIleAspGluPheAspLysMetSerAspAsnAlaArg 615
Db       1875 CTGAGTGACAACGGCATCTGCTGTATCGATGAGTTCGACAAGATGAATGAAAGTACAAGA 1934

Qy        616 SerMetLeuHisGluValMetGluGlnGlnThrValSerValAlaLysGlyGlyIleIle 635
Db       1935 TCGGTATTGCATGAAGTCATGGAACAGCAGACTCTGTCCATTGCAAAGGCTGGGATCATC 1994

Qy        636 AlaSerLeuAsnAlaArgThrSerValLeuAlaCysAlaAsnProSerGlySerArgTyr 655
Db       1995 TGTCAGCTCAATGCGCGCACCTCTGTCCTGGCAGCAGCAAATCCCATTGAGTCTCAGTGG 2054

Qy        656 AsnAlaArgLeuSerValIleAspAsnIleGlnLeuProProThrLeuLeuSerArgPhe 675
Db       2055 AATCCTAAAAAAACAACCATTGAAAACATCCAGCTGCCTCATACTTTATTATCAAGGTTT 2114

Qy        676 AspLeuIleTyrLeuMetLeuAspLysProAspGluGlnAsnAspArgArgLeuAlaArg 695
Db       2115 GATTTGATCTTCCTCATGCTGGACCCTCAGGACGAAGCCTATGACAGGCGTCTGGCTCAC 2174

Qy        696 HisLeuValAlaLeuHisTyrGluAsnTyrGluValSerLysGlnAspAlaLeuAspLeu 715
Db       2175 CACCTGGTCGCACTGTACTACCAGAGCGAGGAGCAGGCAGAGGAGGAGCTCCTGGACATG 2234

Qy        716 GlnThrLeuThrAlaTyrIleThrTyrAlaArgGlnHisValHisProThrLeuSerAsp 735
Db       2235 GCGGTGCTAAAGGACTACATTGCCTACGCGCACAGCACCATCATGCCGCGGCTAAGTGAG 2294

Qy        736 GluAlaAlaGluAspLeuIleAsnGlyTyrValGluMetArgGlnLysGlyAsnPhePro 755
Db       2295 GAAGCCAGCCAGGCTCTCATCGAGGCTTATGTAGACATGAGGAAG-----------ATT 2342

Qy        756 GlySerSerLysLysValIleThrAlaThrProArgGlnLeuGluSerMetIleArgIle 775
Db       2343 GGCAGTAGCCGGGGAATGGTTTCTGCATACCCTCGACAGCTAGAGTCATTAATCCGCTTA 2402

Qy        776 SerGluAlaLeuAlaArgMetArgPheSerGluValValGluLysValAspAlaAlaGlu 795
Db       2403 GCAGAAGCCCATGCTAAAGTAAGATTGTCTAACAAAGTTGAAGCCATTGATGTGGAAGAG 2462

Qy        796 AlaValArgLeuLeuAspValAlaLeuGlnGlnSerAlaThrAspHisAlaThrGlyThr 815
Db       2463 GCCAAACGCCTCCATCGGGAAGCTCTGAAGCAGTCTGCAACTGATCCCCGGACTGGCATC 2522

Qy        816 IleAspMetAspLeuIleThrThrGlyValSerAlaSerGluArgIleArgArgAlaAsn 835
Db       2523 GTGGACATATCTATTCTTACTACGGGGATGAGTGCCACCTCTCGTAAACGGAAAGAAGAA 2582

Qy        836 LeuLeuAlaAlaLeuArgGluLeuIleAlaAspLysIleSerProGlySerSerSerGly 855
Db       2583 TTAGCTGAAGCATTGAAAAAGCTTATT---------TTATCTAAGGGCAAAACACCAGCT 2633

Qy        856 LeuLysThrSerGlnLeuLeuGluAspIleArgSerGlnSerSerValAspValSerLeu 875
Db       2634 CTAAAATACCAGCAACTTTTTGAAGATATTCGGGGACAATCTGACATAGCAATTACTAAA 2693

Qy        876 GlnAspIleLysAsnAlaLeuGlySerLeuGlnGlyGluGlyPheLeuThrValHisGly 895
Db       2694 GATATGTTTGAAGAAGCACTGCGTGCCCTGGCAGATGATGATTTCCTGACAGTGACTGGG 2753

Qy        896 AspIleValLys 899
Db       2754 AAGACCGTGCGC 2765
```

```
RESULT 11
ABL65666
ID   ABL65666 standard; DNA; 3273 BP.
XX
AC   ABL65666;
XX
DT   15-MAY-2002  (first entry)
```

```
XX
DE   Lung cancer related gene sequence SEQ ID NO:4003.
XX
KW   Human; cancer; colon; breast; ovary; oesophagus; kidney; thyroid;
KW   stomach; lung; prostate; pancreas; carcinoma; antitumour; cancerous;
KW   cytostatic; gene therapy; antineoplastic; Wilm's tumour; adenocarcinoma;
KW   gene; ds.
XX
OS   Homo sapiens.
XX
PN   WO200194629-A2.
XX
PD   13-DEC-2001.
XX
PF   30-MAY-2001; 2001WO-US010838.
XX
PR   05-JUN-2000; 2000US-0209473P.
PR   05-JUN-2000; 2000US-0209531P.
PR   18-SEP-2000; 2000US-0233133P.
PR   18-SEP-2000; 2000US-0233617P.
PR   20-SEP-2000; 2000US-0234009P.
PR   20-SEP-2000; 2000US-0234034P.
PR   20-SEP-2000; 2000US-0234052P.
PR   22-SEP-2000; 2000US-0234509P.
PR   22-SEP-2000; 2000US-0234567P.
PR   25-SEP-2000; 2000US-0234923P.
PR   25-SEP-2000; 2000US-0234924P.
PR   25-SEP-2000; 2000US-0235077P.
PR   25-SEP-2000; 2000US-0235082P.
PR   25-SEP-2000; 2000US-0235134P.
PR   25-SEP-2000; 2000US-0235280P.
PR   26-SEP-2000; 2000US-0235637P.
PR   26-SEP-2000; 2000US-0235638P.
PR   27-SEP-2000; 2000US-0235711P.
PR   27-SEP-2000; 2000US-0235720P.
PR   27-SEP-2000; 2000US-0235840P.
PR   27-SEP-2000; 2000US-0235863P.
PR   28-SEP-2000; 2000US-0236028P.
PR   28-SEP-2000; 2000US-0236032P.
PR   28-SEP-2000; 2000US-0236033P.
PR   28-SEP-2000; 2000US-0236034P.
PR   28-SEP-2000; 2000US-0236109P.
PR   28-SEP-2000; 2000US-0236111P.
PR   29-SEP-2000; 2000US-0236842P.
PR   29-SEP-2000; 2000US-0236891P.
PR   02-OCT-2000; 2000US-0237172P.
PR   02-OCT-2000; 2000US-0237173P.
PR   02-OCT-2000; 2000US-0237278P.
PR   02-OCT-2000; 2000US-0237294P.
PR   02-OCT-2000; 2000US-0237295P.
PR   02-OCT-2000; 2000US-0237316P.
PR   03-OCT-2000; 2000US-0237425P.
PR   03-OCT-2000; 2000US-0237598P.
PR   03-OCT-2000; 2000US-0237604P.
PR   03-OCT-2000; 2000US-0237606P.
PR   03-OCT-2000; 2000US-0237608P.
PR   01-NOV-2000; 2000US-0244867P.
PR   01-NOV-2000; 2000US-0245084P.
XX
PA   (AVAL-) AVALON PHARM.
XX
PI   Young PE,  Augustus M,  Carter KC,  Ebner R,  Endress G,  Horrigan S;
PI   Soppet DR,  Weaver Z;
XX
DR   WPI; 2002-188264/24.
XX
PT   Screening for anti-neoplastic agent involves exposing cells to a chemical
PT   agent to be tested for anti-neoplastic activity, and determining a change
PT   in expression of a gene of a signature gene set.
XX
PS   Claim 1; SEQ ID NO 4003; 44pp; English.
XX
CC   The present invention describes a method (M1) for screening for an anti-
CC   neoplastic agent. The method involves exposing cells to a chemical agent
CC   to be tested for anti-neoplastic activity, determining a change in
CC   expression of at least one gene (I) of a signature gene set, where (I)
CC   comprises a sequence (S) selected from 8447 sequences (given in ABL61664
CC   to ABL70110), or is at least 95% identical to (S), where a change in
CC   expression is indicative of anti-neoplastic activity. (I) has cytostatic
CC   activity and can be used in gene therapy. M1 can be used for screening an
CC   anti-neoplastic agent, and can be used for producing a product which is
CC   the data collected with respect to the anti-neoplastic agent as a result
CC   of M1, and the data is sufficient to convey the chemical structure and/or
CC   properties of the agent. M1 can be used in the treatment of cancer such
CC   as colon, breast, stomach, lung, thyroid, oesophageal, ovarian, kidney,
CC   prostate or pancreatic cancer, adenocarcinoma, carcinoma, clear cell
CC   cancer, infiltrating ductal cancer, infiltrating lobular cancer, squamous
CC   cell carcinoma, neuroendocrine carcinoma, papillary carcinoma and Wilm's
CC   tumour
XX
SQ   Sequence 3273 BP; 876 A; 809 C; 859 G; 729 T; 0 U; 0 Other;

Alignment Scores:
Pred. No.:              6.4e-98         Length:         3273
Score:                  1817.50         Matches:        406
Percent Similarity:     57.78%          Conservative:   151
Best Local Similarity:  42.12%          Mismatches:     256
Query Match:            39.43%          Indels:         151
DB:                     6               Gaps:           18

US-10-768-511-8 (1-901) x ABL65666 (1-3273)

Qy         38 AlaAlaSerProValAlaGlyArgArgAlaValArgGlnThrProThrSerAlaValArg 57
              |||   |||       :::||||||                  |||   :::|||   |||
Db         21 GCGACGTCGTCGGCAAGCGGCCGC---------------CTTCCACGTAACGCG---CGC 62

Qy         58 ArgArgGlyArgGluThrAspSerAlaArgArgArgArgSerArg--------------- 72
              ||||||||||||            :::||||||   |||   |||
Db         63 CGGCGGGGAGGGCGTTGGCGCGGAGCCGACGGGAACGTCCGCGCTGCGGAGCAGGGCAG 122

Qy         73 ---SerArgSerLeuGlyAsnSerValTyrSerSerProTyrAspAlaGlyThrProGly 91
                 ||||||      |||         :::   |||      ||||||||||||
Db        123 GGAAGCCGGGAGGCGGGC---CCGGCCCGAGCTTGTCCTTGTCGCGCAGGTACTCCGAGC 179

Qy         92 ThrProGlyThrProValAlaThrProVal--------------------TyrAlaThr 104
              |||      :::|||   :::||||||                       ||||||
Db        180 ACTATGTCGTCCCCGGCGTCGACCCCGAGCCGCCGCGGCAGCCGGCGTGGAAGGGCCACC 239

Qy        105 ProValGlyThrPro--------------------------------------------- 109
              |||      ||||||
Db        240 CCCGCCCAGACGCCTCGGAGTGAGGATGCCAGGTCATCTCCCTCTCAGAGACGTAGAGGC 299

Qy        109 ------------------------------------------------------------ 109

Db        300 GAGGATTCCACCTCCACGGGGGAGTTGCAGCCGATGCCAACCTCGCCTGGAGTGGACCTG 359

Qy        109 ------------------------------------------------------------ 109

Db        360 CAGAGCACTGCTGCGCAGGACGTGCTGTTTTCCAGCCCTCCCCAAATGCATTCTTCAGCT 419

Qy        110 ------------------------------------MetGlyThrProSerPheHis--- 116
                                                  ||||||||||||
Db        420 ATCCCTCTTGACTTTGATGTTAGTTCACCACTGACATACGGCACTCCCAGCTCTCGGGTA 479

Qy        117 ------------------------ArgGlyThrProGlnTyrLysGlnArgSerGluLeu 128
                                      ||||||||||||      :::||||||   :::|||
Db        480 GAGGGAACCCCAAGAAGTGGTGTTAGGGGCACACCT---GTGAGACAGAGGCCTGACCTG 536

Qy        129 GlySerGlnGlyLysProLeuHisArgArgArgArgSerGlnSerArgGluProGlyHis 148
              ||||||      |||   |||               :::|||
Db        537 GGCTCTGCACAGAAGGGCCTGCAAGTGGATCTGCAGTCT--------------------- 575

Qy        149 ArgSerProSerArgGluProSerAlaAspGlyArgProSerGluSerAlaGluProAsp 168
                                          ||||||      :::|||
Db        576 ---------------------------GACGGGGCAGCAGCAGAAGATATAGTGGCAAGT 608
```

```
Qy   169 AspThrLeuGlyGlyGluTyrAlaTyrValTrpGlyThrAsnValAsnIleProAspVal 188
Db   609 GAGCAGTCTCTAGGCCAAAAACTTGTGATCTGGGGAACAGATGTAAATGTGGCAGCATGC 668

Qy   189 LeuArgAlaIleArgArgPheLeuHisAsnTyr-----------------------Arg 200
Db   669 AAAGAAAACTTTCAGAGATTTCTTCAGCGTTTTATTGACCCTCTGGCTAAAGAAGAAGAA 728

Qy   201 SerSerAlaHisAspLeuAsnSerLys---TyrIleGlnIleIleGluGluThrValGlu 219
Db   729 AATGTTGGCATAGATATTACTGAACCTCTATACATGCAACGACTTGGGGAGATTAATGTT 788

Qy   220 ArgGluGluAspThrLeuAsnIleAspMetSerAspIleTyrAspHisAspProAspLeu 239
Db   789 ATTGGTGAGCAATTTTTAAATGTGAACTGTGAACACATCAAATCATTTGACAAAAATTTG 848

Qy   240 TyrAlaLysIleValArgTyrProLeuAspIleIleProLeuLeuAspThrGluCysGln 259
Db   849 TACAGACAACTCATCTCTTACCCACAGGAAGTTATTCCAACTTTTGACATGGCTGTCAAT 908

Qy   260 GluValAlaThrSerLeuLeuPro------ThrPheGluLysHisIleGluAlaArgPro 277
Db   909 GAAATCTTCTTTGACCGTTACCCTGACTCAATCTTAGAACATCAGATTCAAGTAAGACCA 968

Qy   278 PheAsnLeuLysAlaSerValHisMetArgGluLeuAsnProSerAspIleAspLysLeu 297
Db   969 TTCAACGCATTGAAGACTAAGAATATGAGAAACCTGAATCCAGAAGACATTGACCAGCTC 1028

Qy   298 ValSerValLysGlyMetValIleArgCysSerSerIleIleProGluIleLysGlyAla 317
Db  1029 ATCACCATCAGCGGCATGGTGATCAGGACATCCCAGCTGATTCCCGAGATGCAGGAGGCC 1088

Qy   318 PhePheLysCysLeuValCysGlyHisSerProProLeuValThrValValLysGlyArg 337
Db  1089 TTCTTCCAGTGCCAAGTGTGTGCCCACACG---ACCCGGGTGGAGATGGACCGCGGCCGC 1145

Qy   338 ValGluGluProThrArgCysGluLysProGluCysAlaAlaArgAsnAlaMetSerLeu 357
Db  1146 ATTGCAGAGCCCAGTGTGTGCGGGCGC------TGCCACACCACCCACAGCATGGCACTC 1199

Qy   358 IleHisAsnArgCysThrPheAlaAsnLysGlnIleValArgLeuGlnGluThrProAsp 377
Db  1200 ATCCACAACCGCTCCCTCTTCTCTGACAAGCAGATGATCAAGCTTCAGGAGTCTCCGGAA 1259

Qy   378 AlaIleProGluGlyGluThrProHisThrValSerMetCysLeuTyrAsnThrMetVal 397
Db  1260 GACATGCCTGCAGGGCAGACACCACACACAGTTATCCTGTTTGCTCACAATGATCTCGTT 1319

Qy   398 AspAlaValLysProGlyAspArgIleGluValThrGlyValPheLysAlaMetAlaVal 417
Db  1320 GACAAGGTCCAGCCTGGGGACAGAGTGAATGTTACAGGCATCTATCGAGCTGTGCCTATT 1379

Qy   418 ArgValGlyProAsnGlnArgThrLeuArgAlaLeuTyrLysThrTyrIleAspCysVal 437
Db  1380 CGAGTCAATCCAAGAGTGAGTAATGTGAAGTCTGTCTACAAAACCCACATTGATGTCATT 1439

Qy   438 HisValLysLysSerAspArgGlyArgLeuGlnThrGluAspProMetGluMetAspLys 457
Db  1440 CATTATCGGAAAACGGATGCAAAACGTCTGCAT-------------------------- 1472

Qy   458 GluAsnAspMetTyrAlaGlyTyrHisGluSerAspThrSerGluAlaAlaAsnGluAla 477
Db  1473 ------------------GGCCTTGATGAAGAAGCAGAACAGAAACTTTTTTCAGAGAAA 1514

Qy   478 LysIleGlnLysLeuLysGluLeuSerLysLeuProGlyIleTyrAspArgLeuSerArg 497
Db  1515 CGTGTGGAATTGCTTAAGGAACTTTCCAGGAAACCAGACATTTATGAGAGGCTTGCTTCA 1574

Qy   498 SerLeuAlaProSerIleTrpGluLeuGluAspIleLysLysGlyLeuLeuCysGlnLeu 517
Db  1575 GCCTTGGCTCCAAGCATTTATGAACATGAAGATATAAAGAAGGGAATTTTGCTTCAGCTC 1634

Qy   518 PheGlyGlyLysAlaLysLysIle------ProSerGlyAlaSerPheArgGlyAspIle 535
Db  1635 TTTGGCGGGACAAGGAAGGATTTTAGTCACACTGGAAGGGGCAAATTTCGGGCTGAGATC 1694

Qy   536 AsnValLeuLeuValGlyAspProGlyThrSerLysSerGlnLeuLeuGlnTyrValHis 555
Db  1695 AACATCTTGCTGTGTGGCGACCCTGGTACCAGCAAGTCCCAGCTGCTGCAGTACGTGTAC 1754

Qy   556 LysIleAlaProArgGlyIleTyrThrSerGlyArgGlySerSerAlaValGlyLeuThr 575
Db  1755 AACCTCGTCCCCAGGGGCCAGTACACGTCTGGGAAGGGCTCCAGTGCAGTTGGCCTCACT 1814

Qy   576 AlaTyrValThrLysAspProGluThrArgGluThrValLeuGluSerGlyAlaLeuVal 595
Db  1815 GCGTACGTAATGAAAGACCCTGAGACAAGGCAGCTGGTCCTGCAGACAGGTGCTCTTGTC 1874

Qy   596 LeuSerAspArgGlyIleCysCysIleAspGluPheAspLysMetSerAspAsnAlaArg 615
Db  1875 CTGAGTGACAACGGCATCTGCTGTATCGATGAGTTCGACAAGATGAATGAAAGTACAAGA 1934

Qy   616 SerMetLeuHisGluValMetGluGlnGlnThrValSerValAlaLysGlyGlyIleIle 635
Db  1935 TCGGTATTGCATGAAGTCATGGAACAGCAGACTCTGTCCATTGCAAAGGCTGGGATCATC 1994

Qy   636 AlaSerLeuAsnAlaArgThrSerValLeuAlaCysAlaAsnProSerGlySerArgTyr 655
Db  1995 TGTCAGCTCAATGCGCGCACCTCTGTCCTGGCAGCAGCAAATCCCATTGAGTCTCAGTGG 2054

Qy   656 AsnAlaArgLeuSerValIleAspAsnIleGlnLeuProProThrLeuLeuSerArgPhe 675
Db  2055 AATCCTAAAAAAACAACCATTGAAAACATCCAGCTGCCTCATACTTTATTATCAAGGTTT 2114

Qy   676 AspLeuIleTyrLeuMetLeuAspLysProAspGluGlnAsnAspArgArgLeuAlaArg 695
Db  2115 GATTTGATCTTCCTCATGCTGGACCCTCAGGACGAAGCCTATGACAGGCGTCTGGCTCAC 2174

Qy   696 HisLeuValAlaLeuHisTyrGluAsnTyrGluValSerLysGlnAspAlaLeuAspLeu 715
Db  2175 CACCTGGTCGCACTGTACTACCAGAGCGAGGAGCAGGCAGAGGAGGAGCTCCTGGACATG 2234

Qy   716 GlnThrLeuThrAlaTyrIleThrTyrAlaArgGlnHisValHisProThrLeuSerAsp 735
Db  2235 GCGGTGCTAAAGGACTACATTGCCTACGCGCACAGCACCATCATGCCGCGGCTAAGTGAG 2294

Qy   736 GluAlaAlaGluAspLeuIleAsnGlyTyrValGluMetArgGlnLysGlyAsnPhePro 755
Db  2295 GAAGCCAGCCAGGCTCTCATCGAGGCTTATGTAGACATGAGGAAG-----------ATT 2342

Qy   756 GlySerSerLysLysValIleThrAlaThrProArgGlnLeuGluSerMetIleArgIle 775
Db  2343 GGCAGTAGCCGGGGAATGGTTTCTGCATACCCTCGACAGCTAGAGTCATTAATCCGCTTA 2402

Qy   776 SerGluAlaLeuAlaArgMetArgPheSerGluValValGluLysValAspAlaAlaGlu 795
Db  2403 GCAGAAGCCCATGCTAAAGTAAGATTGTCTAACAAAGTTGAAGCCATTGATGTGGAAGAG 2462

Qy   796 AlaValArgLeuLeuAspValAlaLeuGlnGlnSerAlaThrAspHisAlaThrGlyThr 815
Db  2463 GCCAAACGCCTCCATCGGGAAGCTCTGAAGCAGTCTGCAACTGATCCCCGGACTGGCATC 2522

Qy   816 IleAspMetAspLeuIleThrThrGlyValSerAlaSerGluArgIleArgArgAlaAsn 835
Db  2523 GTGGACATATCTATTCTTACTACGGGGATGAGTGCCACCTCTCGTAAACGGAAAGAAGAA 2582

Qy   836 LeuLeuAlaAlaLeuArgGluLeuIleAlaAspLysIleSerProGlySerSerSerGly 855
Db  2583 TTAGCTGAAGCATTGAAAAAGCTTATT---------TTATCTAAGGGCAAAACACCAGCT 2633

Qy   856 LeuLysThrSerGlnLeuLeuGluAspIleArgSerGlnSerSerValAspValSerLeu 875
Db  2634 CTAAAATACCAGCAACTTTTTGAAGATATTCGGGGACAATCTGACATAGCAATTACTAAA 2693

Qy   876 GlnAspIleLysAsnAlaLeuGlySerLeuGlnGlyGluGlyPheLeuThrValHisGly 895
```

```
Db       2694 GATATGTTTGAAGAAGCACTGCGTGCCCTGGCAGATGATGATTTCCTGACAGTGACTGGG 2753

Qy        896 AspIleValLys 899
              |||:::
Db       2754 AAGACCGTGCGC 2765

RESULT 12
ABL65259
ID   ABL65259 standard; DNA; 3273 BP.
XX
AC   ABL65259;
XX
DT   15-MAY-2002  (first entry)
XX
DE   Lung cancer related gene sequence SEQ ID NO:3596.
XX
KW   Human; cancer; colon; breast; ovary; oesophagus; kidney; thyroid;
KW   stomach; lung; prostate; pancreas; carcinoma; antitumour; cancerous;
KW   cytostatic; gene therapy; antineoplastic; Wilm's tumour; adenocarcinoma;
KW   gene; ds.
XX
OS   Homo sapiens.
XX
PN   WO200194629-A2.
XX
PD   13-DEC-2001.
XX
PF   30-MAY-2001; 2001WO-US010838.
XX
PR   05-JUN-2000; 2000US-0209473P.
PR   05-JUN-2000; 2000US-0209531P.
PR   18-SEP-2000; 2000US-0233133P.
PR   18-SEP-2000; 2000US-0233617P.
PR   20-SEP-2000; 2000US-0234009P.
PR   20-SEP-2000; 2000US-0234034P.
PR   20-SEP-2000; 2000US-0234052P.
PR   22-SEP-2000; 2000US-0234509P.
PR   22-SEP-2000; 2000US-0234567P.
PR   25-SEP-2000; 2000US-0234923P.
PR   25-SEP-2000; 2000US-0234924P.
PR   25-SEP-2000; 2000US-0235077P.
PR   25-SEP-2000; 2000US-0235082P.
PR   25-SEP-2000; 2000US-0235134P.
PR   25-SEP-2000; 2000US-0235280P.
PR   26-SEP-2000; 2000US-0235637P.
PR   26-SEP-2000; 2000US-0235638P.
PR   27-SEP-2000; 2000US-0235711P.
PR   27-SEP-2000; 2000US-0235720P.
PR   27-SEP-2000; 2000US-0235840P.
PR   27-SEP-2000; 2000US-0235863P.
PR   28-SEP-2000; 2000US-0236028P.
PR   28-SEP-2000; 2000US-0236032P.
PR   28-SEP-2000; 2000US-0236033P.
PR   28-SEP-2000; 2000US-0236034P.
PR   28-SEP-2000; 2000US-0236109P.
PR   28-SEP-2000; 2000US-0236111P.
PR   29-SEP-2000; 2000US-0236842P.
PR   29-SEP-2000; 2000US-0236891P.
PR   02-OCT-2000; 2000US-0237172P.
PR   02-OCT-2000; 2000US-0237173P.
PR   02-OCT-2000; 2000US-0237278P.
PR   02-OCT-2000; 2000US-0237294P.
PR   02-OCT-2000; 2000US-0237295P.
PR   02-OCT-2000; 2000US-0237316P.
PR   03-OCT-2000; 2000US-0237425P.
PR   03-OCT-2000; 2000US-0237598P.
PR   03-OCT-2000; 2000US-0237604P.
PR   03-OCT-2000; 2000US-0237606P.
PR   03-OCT-2000; 2000US-0237608P.
PR   01-NOV-2000; 2000US-0244867P.
PR   01-NOV-2000; 2000US-0245084P.
XX
PA   (AVAL-) AVALON PHARM.
XX
PI   Young PE,  Augustus M,  Carter KC,  Ebner R,  Endress G,  Horrigan S;
PI   Soppet DR,  Weaver Z;
XX
DR   WPI; 2002-188264/24.
XX
PT   Screening for anti-neoplastic agent involves exposing cells to a chemical
PT   agent to be tested for anti-neoplastic activity, and determining a change
PT   in expression of a gene of a signature gene set.
XX
PS   Claim 1; SEQ ID NO 3596; 44pp; English.
XX
CC   The present invention describes a method (M1) for screening for an anti-
CC   neoplastic agent. The method involves exposing cells to a chemical agent
CC   to be tested for anti-neoplastic activity, determining a change in
CC   expression of at least one gene (I) of a signature gene set, where (I)
CC   comprises a sequence (S) selected from 8447 sequences (given in ABL61664
CC   to ABL70110), or is at least 95% identical to (S), where a change in
CC   expression is indicative of anti-neoplastic activity. (I) has cytostatic
CC   activity and can be used in gene therapy. M1 can be used for screening an
CC   anti-neoplastic agent, and can be used for producing a product which is
CC   the data collected with respect to the anti-neoplastic agent as a result
CC   of M1, and the data is sufficient to convey the chemical structure and/or
CC   properties of the agent. M1 can be used in the treatment of cancer such
CC   as colon, breast, stomach, lung, thyroid, oesophageal, ovarian, kidney,
CC   prostate or pancreatic cancer, adenocarcinoma, carcinoma, clear cell
CC   cancer, infiltrating ductal cancer, infiltrating lobular cancer, squamous
CC   cell carcinoma, neuroendocrine carcinoma, papillary carcinoma and Wilm's
CC   tumour
XX
SQ   Sequence 3273 BP; 876 A; 809 C; 859 G; 729 T; 0 U; 0 Other;

Alignment Scores:
Pred. No.:              6.4e-98       Length:        3273
Score:                  1817.50       Matches:       406
Percent Similarity:     57.78%        Conservative:  151
Best Local Similarity:  42.12%        Mismatches:    256
Query Match:            39.43%        Indels:        151
DB:                     6             Gaps:          18

US-10-768-511-8 (1-901) x ABL65259 (1-3273)

Qy         38 AlaAlaSerProValAlaGlyArgArgAlaValArgGlnThrProThrSerAlaValArg 57
              |||   |||      :::||||||                 |||   :::|||   |||
Db         21 GCGACGTCGTCGGCAAGCGGCCGC---------------CTTCCACGTAACGCG---CGC 62

Qy         58 ArgArgGlyArgGluThrAspSerAlaArgArgArgArgSerArg--------------- 72
              ||||||||||||            :::||||||   |||   |||
Db         63 CGGCGGGGGAGGGCGTTGGCGCGGAGCCGACGGGAACGTCCGCGCTGCGGAGCAGGGCAG 122

Qy         73 ---SerArgSerLeuGlyAsnSerValTyrSerSerProTyrAspAlaGlyThrProGly 91
                 ||||||      |||         :::   |||      ||||||||||||
Db        123 GGAAGCCGGGAGGCGGGC---CCGGCCCGAGCTTGTCCTTGTCGCGCAGGTACTCCGAGC 179

Qy         92 ThrProGlyThrProValAlaThrProVal---------------------TyrAlaThr 104
              |||   :::|||   :::||||||                              ||||||
Db        180 ACTATGTCGTCCCCGGCGTCGACCCCGAGCCGCCGCGGCAGCCGGCGTGGAAGGGCCACC 239

Qy        105 ProValGlyThrPro--------------------------------------------- 109
              |||   ||||||
Db        240 CCCGCCCAGACGCCTCGGAGTGAGGATGCCAGGTCATCTCCCTCTCAGAGACGTAGAGGC 299

Qy        109 ------------------------------------------------------------ 109

Db        300 GAGGATTCCACCTCCACGGGGGAGTTGCAGCCGATGCCAACCTCGCCTGGAGTGGACCTG 359

Qy        109 ------------------------------------------------------------ 109

Db        360 CAGAGCACTGCTGCGCAGGACGTGCTGTTTTCCAGCCCTCCCCAAATGCATTCTTCAGCT 419

Qy        110 ------------------------------------MetGlyThrProSerPheHis--- 116
                                                  ||||||||||||
Db        420 ATCCCTCTTGACTTTGATGTTAGTTCACCACTGACATACGGCACTCCCAGCTCTCGGGTA 479
```

```
Qy    117 ------------------------ArgGlyThrProGlnTyrLysGlnArgSerGluLeu 128
Db    480 GAGGGAACCCCAAGAAGTGGTGTTAGGGGCACACCT---GTGAGACAGAGGCCTGACCTG 536

Qy    129 GlySerGlnGlyLysProLeuHisArgArgArgArgSerGlnSerArgGluProGlyHis 148
Db    537 GGCTCTGCACAGAAGGGCCTGCAAGTGGATCTGCAGTCT--------------------- 575

Qy    149 ArgSerProSerArgGluProSerAlaAspGlyArgProSerGluSerAlaGluProAsp 168
Db    576 --------------------------GACGGGGCAGCAGCAGAAGATATAGTGGCAAGT 608

Qy    169 AspThrLeuGlyGlyGluTyrAlaTyrValTrpGlyThrAsnValAsnIleProAspVal 188
Db    609 GAGCAGTCTCTAGGCCAAAAACTTGTGATCTGGGGAACAGATGTAAATGTGGCAGCATGC 668

Qy    189 LeuArgAlaIleArgArgPheLeuHisAsnTyr-----------------------Arg 200
Db    669 AAAGAAAACTTTCAGAGATTTCTTCAGCGTTTTATTGACCCTCTGGCTAAAGAAGAAGAA 728

Qy    201 SerSerAlaHisAspLeuAsnSerLys---TyrIleGlnIleIleGluGluThrValGlu 219
Db    729 AATGTTGGCATAGATATTACTGAACCTCTATACATGCAACGACTTGGGGAGATTAATGTT 788

Qy    220 ArgGluGluAspThrLeuAsnIleAspMetSerAspIleTyrAspHisAspProAspLeu 239
Db    789 ATTGGTGAGCAATTTTTAAATGTGAACTGTGAACACATCAAATCATTTGACAAAAATTTG 848

Qy    240 TyrAlaLysIleValArgTyrProLeuAspIleIleProLeuLeuAspThrGluCysGln 259
Db    849 TACAGACAACTCATCTCTTACCCACAGGAAGTTATTCCAACTTTTGACATGGCTGTCAAT 908

Qy    260 GluValAlaThrSerLeuLeuPro------ThrPheGluLysHisIleGluAlaArgPro 277
Db    909 GAAATCTTCTTTGACCGTTACCCTGACTCAATCTTAGAACATCAGATTCAAGTAAGACCA 968

Qy    278 PheAsnLeuLysAlaSerValHisMetArgGluLeuAsnProSerAspIleAspLysLeu 297
Db    969 TTCAACGCATTGAAGACTAAGAATATGAGAAACCTGAATCCAGAAGACATTGACCAGCTC 1028

Qy    298 ValSerValLysGlyMetValIleArgCysSerSerIleIleProGluIleLysGlyAla 317
Db   1029 ATCACCATCAGCGGCATGGTGATCAGGACATCCCAGCTGATTCCCGAGATGCAGGAGGCC 1088

Qy    318 PhePheLysCysLeuValCysGlyHisSerProProLeuValThrValValLysGlyArg 337
Db   1089 TTCTTCCAGTGCCAAGTGTGTGCCCACACG---ACCCGGGTGGAGATGGACCGCGGCCGC 1145

Qy    338 ValGluGluProThrArgCysGluLysProGluCysAlaAlaArgAsnAlaMetSerLeu 357
Db   1146 ATTGCAGAGCCCAGTGTGTGCGGGCGC------TGCCACACCACCCACAGCATGGCACTC 1199

Qy    358 IleHisAsnArgCysThrPheAlaAsnLysGlnIleValArgLeuGlnGluThrProAsp 377
Db   1200 ATCCACAACCGCTCCCTCTTCTCTGACAAGCAGATGATCAAGCTTCAGGAGTCTCCGGAA 1259

Qy    378 AlaIleProGluGlyGluThrProHisThrValSerMetCysLeuTyrAsnThrMetVal 397
Db   1260 GACATGCCTGCAGGGCAGACACCACACACAGTTATCCTGTTTGCTCACAATGATCTCGTT 1319

Qy    398 AspAlaValLysProGlyAspArgIleGluValThrGlyValPheLysAlaMetAlaVal 417
Db   1320 GACAAGGTCCAGCCTGGGGACAGAGTGAATGTTACAGGCATCTATCGAGCTGTGCCTATT 1379

Qy    418 ArgValGlyProAsnGlnArgThrLeuArgAlaLeuTyrLysThrTyrIleAspCysVal 437
Db   1380 CGAGTCAATCCAAGAGTGAGTAATGTGAAGTCTGTCTACAAAACCCACATTGATGTCATT 1439

Qy    438 HisValLysLysSerAspArgGlyArgLeuGlnThrGluAspProMetGluMetAspLys 457
Db   1440 CATTATCGGAAAACGGATGCAAAACGTCTGCAT-------------------------- 1472

Qy    458 GluAsnAspMetTyrAlaGlyTyrHisGluSerAspThrSerGluAlaAlaAsnGluAla 477
Db   1473 ------------------GGCCTTGATGAAGAAGCAGAACAGAAACTTTTTCAGAGAAA 1514

Qy    478 LysIleGlnLysLeuLysGluLeuSerLysLeuProGlyIleTyrAspArgLeuSerArg 497
Db   1515 CGTGTGGAATTGCTTAAGGAACTTTCCAGGAAACCAGACATTTATGAGAGGCTTGCTTCA 1574

Qy    498 SerLeuAlaProSerIleTrpGluLeuGluAspIleLysLysGlyLeuLeuCysGlnLeu 517
Db   1575 GCCTTGGCTCCAAGCATTTATGAACATGAAGATATAAAGAAGGGAATTTTGCTTCAGCTC 1634

Qy    518 PheGlyGlyLysAlaLysLysIle------ProSerGlyAlaSerPheArgGlyAspIle 535
Db   1635 TTTGGCGGGACAAGGAAGGATTTTAGTCACACTGGAAGGGGCAAATTTCGGGCTGAGATC 1694

Qy    536 AsnValLeuLeuValGlyAspProGlyThrSerLysSerGlnLeuLeuGlnTyrValHis 555
Db   1695 AACATCTTGCTGTGTGGCGACCCTGGTACCAGCAAGTCCCAGCTGCTGCAGTACGTGTAC 1754

Qy    556 LysIleAlaProArgGlyIleTyrThrSerGlyArgGlySerSerAlaValGlyLeuThr 575
Db   1755 AACCTCGTCCCCAGGGGCCAGTACACGTCTGGGAAGGGCTCCAGTGCAGTTGGCCTCACT 1814

Qy    576 AlaTyrValThrLysAspProGluThrArgGluThrValLeuGluSerGlyAlaLeuVal 595
Db   1815 GCGTACGTAATGAAAGACCCTGAGACAAGGCAGCTGGTCCTGCAGACAGGTGCTCTTGTC 1874

Qy    596 LeuSerAspArgGlyIleCysCysIleAspGluPheAspLysMetSerAspAsnAlaArg 615
Db   1875 CTGAGTGACAACGGCATCTGCTGTATCGATGAGTTCGACAAGATGAATGAAAGTACAAGA 1934

Qy    616 SerMetLeuHisGluValMetGluGlnGlnThrValSerValAlaLysGlyGlyIleIle 635
Db   1935 TCGGTATTGCATGAAGTCATGGAACAGCAGACTCTGTCCATTGCAAAGGCTGGGATCATC 1994

Qy    636 AlaSerLeuAsnAlaArgThrSerValLeuAlaCysAlaAsnProSerGlySerArgTyr 655
Db   1995 TGTCAGCTCAATGCGCGCACCTCTGTCCTGGCAGCAGCAAATCCCATTGAGTCTCAGTGG 2054

Qy    656 AsnAlaArgLeuSerValIleAspAsnIleGlnLeuProProThrLeuLeuSerArgPhe 675
Db   2055 AATCCTAAAAAAACAACCATTGAAAACATCCAGCTGCCTCATACTTTATTATCAAGGTTT 2114

Qy    676 AspLeuIleTyrLeuMetLeuAspLysProAspGluGlnAsnAspArgArgLeuAlaArg 695
Db   2115 GATTTGATCTTCCTCATGCTGGACCCTCAGGACGAAGCCTATGACAGGCGTCTGGCTCAC 2174

Qy    696 HisLeuValAlaLeuHisTyrGluAsnTyrGluValSerLysGlnAspAlaLeuAspLeu 715
Db   2175 CACCTGGTCGCACTGTACTACCAGAGCGAGGAGCAGGCAGAGGAGGAGCTCCTGGACATG 2234

Qy    716 GlnThrLeuThrAlaTyrIleThrTyrAlaArgGlnHisValHisProThrLeuSerAsp 735
Db   2235 GCGGTGCTAAAGGACTACATTGCCTACGCGCACAGCACCATCATGCCGCGGCTAAGTGAG 2294

Qy    736 GluAlaAlaGluAspLeuIleAsnGlyTyrValGluMetArgGlnLysGlyAsnPhePro 755
Db   2295 GAAGCCAGCCAGGCTCTCATCGAGGCTTATGTAGACATGAGGAAG-----------ATT 2342

Qy    756 GlySerSerLysLysValIleThrAlaThrProArgGlnLeuGluSerMetIleArgIle 775
Db   2343 GGCAGTAGCCGGGGAATGGTTTCTGCATACCCTCGACAGCTAGAGTCATTAATCCGCTTA 2402

Qy    776 SerGluAlaLeuAlaArgMetArgPheSerGluValValGluLysValAspAlaAlaGlu 795
Db   2403 GCAGAAGCCCATGCTAAAGTAAGATTGTCTAACAAAGTTGAAGCCATTGATGTGGAAGAG 2462

Qy    796 AlaValArgLeuLeuAspValAlaLeuGlnGlnSerAlaThrAspHisAlaThrGlyThr 815
Db   2463 GCCAAACGCCTCCATCGGGAAGCTCTGAAGCAGTCTGCAACTGATCCCCGGACTGGCATC 2522

Qy    816 IleAspMetAspLeuIleThrThrGlyValSerAlaSerGluArgIleArgArgAlaAsn 835
```

```
                :::|||:::   ::::::||||||||:::||||||:::   |||   |||:::
Db       2523 GTGGACATATCTATTCTTACTACGGGGATGAGTGCCACCTCTCGTAAACGGAAAGAAGAA 2582

Qy        836 LeuLeuAlaAlaLeuArgGluLeuIleAlaAspLysIleSerProGlySerSerSerGly 855
              |||       ||||||::::::||||||         :::|||   |||   :::
Db       2583 TTAGCTGAAGCATTGAAAAAGCTTATT---------TTATCTAAGGGCAAAACACCAGCT 2633

Qy        856 LeuLysThrSerGlnLeuLeuGluAspIleArgSerGlnSerSerValAspValSerLeu 875
              ||||||      ||||||   |||||||||||   ||||||   :::   ::::::
Db       2634 CTAAAATACCAGCAACTTTTTGAAGATATTCGGGACAATCTGACATAGCAATTACTAAA 2693

Qy        876 GlnAspIleLysAsnAlaLeuGlySerLeuGlnGlyGluGlyPheLeuThrValHisGly 895
                 :::   ||||||   :::|||      :::   ||||||||||||   |||
Db       2694 GATATGTTTGAAGAAGCACTGCGTGCCCTGGCAGATGATGATTTCCTGACAGTGACTGGG 2753

Qy        896 AspIleValLys 899
                    |||:::
Db       2754 AAGACCGTGCGC 2765

RESULT 13
AD019849
ID   AD019849 standard; DNA; 3273 BP.
XX
AC   AD019849;
XX
DT   12-AUG-2004  (first entry)
XX
DE   Human PRO polynucleotide #727.
XX
KW   Human; PRO; gene; ss; immune related disorder;
KW   systemic lupus erythematosus; rheumatoid arthritis; osteoarthritis;
KW   juvenile chronic arthritis; systemic sclerosis; Sjogren's syndrome;
KW   vasculitis; sarcoidosis; autoimmune haemolytic anaemia;
KW   autoimmune thrombocytopenia; thyroiditis; diabetes mellitus;
KW   renal disease; demyelinating disease; central nervous system;
KW   peripheral nervous system; demyelinating polyneuropathy;
KW   Guillain-Barre syndrome;
KW   chronic inflammatory demyelinating polyneuropathy.
XX
OS   Homo sapiens.
XX
PN   WO2004043361-A2.
XX
PD   27-MAY-2004.
XX
PF   06-NOV-2003; 2003WO-US035268.
XX
PR   08-NOV-2002; 2002US-0425235P.
XX
PA   (GETH ) GENENTECH INC.
XX
PI   Fong S,  Dennis K,  Clark H,  Chiu H,  Schoenfeld J,  Williams PM;
PI   Wood WI,  Wu TD;
XX
DR   WPI; 2004-420067/39.
XX
PT   Novel PRO polypeptide e.g., PRO69614, PRO71106, or PRO86388 useful for
PT   treating an immune related disorder such as systemic lupus erythematosus,
PT   rheumatoid arthritis, osteoarthritis, juvenile chronic arthritis or
PT   spondyloarthropathy.
XX
PS   Claim 1; SEQ ID NO 798; 1731pp; English.
XX
CC   The invention relates to human PRO polypeptides and the polynucleotides
CC   encoding them. The polypeptides and polynucleotides are useful for
CC   treating and diagnosing immune related disorders in mammals. The immune
CC   related disorders include systemic lupus erythematosus, rheumatoid
CC   arthritis, osteoarthritis, juvenile chronic arthritis, systemic
CC   sclerosis, Sjogren's syndrome, vasculitis, sarcoidosis, autoimmune
CC   haemolytic anaemia, autoimmune thrombocytopenia, thyroiditis, diabetes
CC   mellitus, immune-mediated renal disease, demyelinating diseases of the
CC   central or peripheral nervous system, demyelinating polyneuropathy,
CC   Guillain-Barre syndrome and chronic inflammatory demyelinating
CC   polyneuropathy. This sequence represents a human PRO polynucleotide of
CC   the invention.
XX
SQ   Sequence 3273 BP; 876 A; 809 C; 859 G; 729 T; 0 U; 0 Other;

Alignment Scores:
Pred. No.:              6.4e-98       Length:         3273
Score:                  1817.50       Matches:        406
Percent Similarity:     57.78%        Conservative:   151
Best Local Similarity:  42.12%        Mismatches:     256
Query Match:            39.43%        Indels:         151
DB:                     12            Gaps:           18

US-10-768-511-8 (1-901) x AD019849 (1-3273)

Qy         38 AlaAlaSerProValAlaGlyArgArgAlaValArgGlnThrProThrSerAlaValArg 57
              |||   |||      :::||||||                  |||   :::|||   |||
Db         21 GCGACGTCGTCGGCAAGCGGCCGC---------------CTTCCACGTAACGCG---CGC 62

Qy         58 ArgArgGlyArgGluThrAspSerAlaArgArgArgArgSerArg--------------- 72
              ||||||||||||            :::||||||   |||   |||
Db         63 CGGCGGGGGAGGGCGTTGGCGCGGAGCCGACGGGAACGTCCGCGCTGCGGAGCAGGGCAG 122

Qy         73 ---SerArgSerLeuGlyAsnSerValTyrSerSerProTyrAspAlaGlyThrProGly 91
                 ||||||      |||            :::   |||      ||||||||||||
Db        123 GGAAGCCGGGAGGCGGGC---CCGGCCCGAGCTTGTCCTTGTCGCGCAGGTACTCCGAGC 179

Qy         92 ThrProGlyThrProValAlaThrProVal--------------------TyrAlaThr 104
              |||      :::|||   :::||||||                        ||||||
Db        180 ACTATGTCGTCCCCGGCGTCGACCCCGAGCCGCCGCGGCAGCCGGCGTGGAAGGGCCACC 239

Qy        105 ProValGlyThrPro--------------------------------------------- 109
              |||      ||||||
Db        240 CCCGCCCAGACGCCTCGGAGTGAGGATGCCAGGTCATCTCCCTCTCAGAGACGTAGAGGC 299

Qy        109 ------------------------------------------------------------ 109

Db        300 GAGGATTCCACCTCCACGGGGGAGTTGCAGCCGATGCCAACCTCGCCTGGAGTGGACCTG 359

Qy        109 ------------------------------------------------------------ 109

Db        360 CAGAGCACTGCTGCGCAGGACGTGCTGTTTTCCAGCCCTCCCCAAATGCATTCTTCAGCT 419

Qy        110 ------------------------------------MetGlyThrProSerPheHis--- 116
                                                  ||||||||||||
Db        420 ATCCCTCTTGACTTTGATGTTAGTTCACCACTGACATACGGCACTCCCAGCTCTCGGGTA 479

Qy        117 -----------------------ArgGlyThrProGlnTyrLysGlnArgSerGluLeu 128
                                     ||||||||||||      :::||||||   :::|||
Db        480 GAGGGAACCCCAAGAAGTGGTGTTAGGGGCACACCT---GTGAGACAGAGGCCTGACCTG 536

Qy        129 GlySerGlnGlyLysProLeuHisArgArgArgArgSerGlnSerArgGluProGlyHis 148
              ||||||      |||   |||            :::|||
Db        537 GGCTCTGCACAGAAGGGCCTGCAAGTGGATCTGCAGTCT--------------------- 575

Qy        149 ArgSerProSerArgGluProSerAlaAspGlyArgProSerGluSerAlaGluProAsp 168
                                         ||||||      :::|||
Db        576 ---------------------------GACGGGGCAGCAGCAGAAGATATAGTGGCAAGT 608

Qy        169 AspThrLeuGlyGlyGluTyrAlaTyrValTrpGlyThrAsnValAsnIleProAspVal 188
              :::         |||:::         :::|||||||||:::||||||:::
Db        609 GAGCAGTCTCTAGGCCAAAAACTTGTGATCTGGGGAACAGATGTAAATGTGGCAGCATGC 668

Qy        189 LeuArgAlaIleArgArgPheLeuHisAsnTyr-----------------------Arg 200
                       :::|||||||||          :::
Db        669 AAAGAAAACTTTCAGAGATTTCTTCAGCGTTTTATTGACCCTCTGGCTAAAGAAGAAGAA 728

Qy        201 SerSerAlaHisAspLeuAsnSerLys---TyrIleGlnIleIleGluGluThrValGlu 219
              :::         |||:::            |||:::|||   :::   |||
Db        729 AATGTTGGCATAGATATTACTGAACCTCTATACATGCAACGACTTGGGGAGATTAATGTT 788

Qy        220 ArgGluGluAspThrLeuAsnIleAspMetSerAspIleTyrAspHisAspProAspLeu 239
                 |||      ||||||::::::            |||         |||   :::|||
```

```
Db    789 ATTGGTGAGCAATTTTTAAATGTGAACTGTGAACACATCAAATCATTTGACAAAAATTTG 848

Qy    240 TyrAlaLysIleValArgTyrProLeuAspIleIleProLeuLeuAspThrGluCysGln 259
Db    849 TACAGACAACTCATCTCTTACCCACAGGAAGTTATTCCAACTTTTGACATGGCTGTCAAT 908

Qy    260 GluValAlaThrSerLeuLeuPro------ThrPheGluLysHisIleGluAlaArgPro 277
Db    909 GAAATCTTCTTTGACCGTTACCCTGACTCAATCTTAGAACATCAGATTCAAGTAAGACCA 968

Qy    278 PheAsnLeuLysAlaSerValHisMetArgGluLeuAsnProSerAspIleAspLysLeu 297
Db    969 TTCAACGCATTGAAGACTAAGAATATGAGAAACCTGAATCCAGAAGACATTGACCAGCTC 1028

Qy    298 ValSerValLysGlyMetValIleArgCysSerSerIleIleProGluIleLysGlyAla 317
Db   1029 ATCACCATCAGCGGCATGGTGATCAGGACATCCCAGCTGATTCCCGAGATGCAGGAGGCC 1088

Qy    318 PhePheLysCysLeuValCysGlyHisSerProProLeuValThrValValLysGlyArg 337
Db   1089 TTCTTCCAGTGCCAAGTGTGTGCCCACACG---ACCCGGGTGGAGATGGACCGCGGCCGC 1145

Qy    338 ValGluGluProThrArgCysGluLysProGluCysAlaAlaArgAsnAlaMetSerLeu 357
Db   1146 ATTGCAGAGCCCAGTGTGTGCGGGCGC------TGCCACACCACCCACAGCATGGCACTC 1199

Qy    358 IleHisAsnArgCysThrPheAlaAsnLysGlnIleValArgLeuGlnGluThrProAsp 377
Db   1200 ATCCACAACCGCTCCCTCTTCTCTGACAAGCAGATGATCAAGCTTCAGGAGTCTCCGGAA 1259

Qy    378 AlaIleProGluGlyGluThrProHisThrValSerMetCysLeuTyrAsnThrMetVal 397
Db   1260 GACATGCCTGCAGGGCAGACACCACACACAGTTATCCTGTTTGCTCACAATGATCTCGTT 1319

Qy    398 AspAlaValLysProGlyAspArgIleGluValThrGlyValPheLysAlaMetAlaVal 417
Db   1320 GACAAGGTCCAGCCTGGGGACAGAGTGAATGTTACAGGCATCTATCGAGCTGTGCCTATT 1379

Qy    418 ArgValGlyProAsnGlnArgThrLeuArgAlaLeuTyrLysThrTyrIleAspCysVal 437
Db   1380 CGAGTCAATCCAAGAGTGAGTAATGTGAAGTCTGTCTACAAAACCCACATTGATGTCATT 1439

Qy    438 HisValLysLysSerAspArgGlyArgLeuGlnThrGluAspProMetGluMetAspLys 457
Db   1440 CATTATCGGAAAACGGATGCAAAACGTCTGCAT------------------------- 1472

Qy    458 GluAsnAspMetTyrAlaGlyTyrHisGluSerAspThrSerGluAlaAlaAsnGluAla 477
Db   1473 ------------------GGCCTTGATGAAGAAGCAGAACAGAAACTTTTTCAGAGAAA 1514

Qy    478 LysIleGlnLysLeuLysGluLeuSerLysLeuProGlyIleTyrAspArgLeuSerArg 497
Db   1515 CGTGTGGAATTGCTTAAGGAACTTTCCAGGAAACCAGACATTTATGAGAGGCTTGCTTCA 1574

Qy    498 SerLeuAlaProSerIleTrpGluLeuGluAspIleLysLysGlyLeuLeuCysGlnLeu 517
Db   1575 GCCTTGGCTCCAAGCATTTATGAACATGAAGATATAAAGAAGGGAATTTTGCTTCAGCTC 1634

Qy    518 PheGlyGlyLysAlaLysLysIle------ProSerGlyAlaSerPheArgGlyAspIle 535
Db   1635 TTTGGCGGGACAAGGAAGGATTTAGTCACACTGGAAGGGGCAAATTTCGGGCTGAGATC 1694

Qy    536 AsnValLeuLeuValGlyAspProGlyThrSerLysSerGlnLeuLeuGlnTyrValHis 555
Db   1695 AACATCTTGCTGTGTGGCGACCCTGGTACCAGCAAGTCCCAGCTGCTGCAGTACGTGTAC 1754

Qy    556 LysIleAlaProArgGlyIleTyrThrSerGlyArgGlySerSerAlaValGlyLeuThr 575
Db   1755 AACCTCGTCCCCAGGGGCCAGTACACGTCTGGGAAGGGCTCCAGTGCAGTTGGCCTCACT 1814

Qy    576 AlaTyrValThrLysAspProGluThrArgGluThrValLeuGluSerGlyAlaLeuVal 595
Db   1815 GCGTACGTAATGAAAGACCCTGAGACAAGGCAGCTGGTCCTGCAGACAGGTGCTCTTGTC 1874

Qy    596 LeuSerAspArgGlyIleCysCysIleAspGluPheAspLysMetSerAspAsnAlaArg 615
Db   1875 CTGAGTGACAACGGCATCTGCTGTATCGATGAGTTCGACAAGATGAATGAAAGTACAAGA 1934

Qy    616 SerMetLeuHisGluValMetGluGlnGlnThrValSerValAlaLysGlyGlyIleIle 635
Db   1935 TCGGTATTGCATGAAGTCATGGAACAGCAGACTCTGTCCATTGCAAAGGCTGGGATCATC 1994

Qy    636 AlaSerLeuAsnAlaArgThrSerValLeuAlaCysAlaAsnProSerGlySerArgTyr 655
Db   1995 TGTCAGCTCAATGCGCGCACCTCTGTCCTGGCAGCAGCAAATCCCATTGAGTCTCAGTGG 2054

Qy    656 AsnAlaArgLeuSerValIleAspAsnIleGlnLeuProProThrLeuLeuSerArgPhe 675
Db   2055 AATCCTAAAAAAACAACCATTGAAAACATCCAGCTGCCTCATACTTTATTATCAAGGTTT 2114

Qy    676 AspLeuIleTyrLeuMetLeuAspLysProAspGluGlnAsnAspArgArgLeuAlaArg 695
Db   2115 GATTTGATCTTCCTCATGCTGGACCCTCAGGACGAAGCCTATGACAGGCGTCTGGCTCAC 2174

Qy    696 HisLeuValAlaLeuHisTyrGluAsnTyrGluValSerLysGlnAspAlaLeuAspLeu 715
Db   2175 CACCTGGTCGCACTGTACTACCAGAGCGAGGAGCAGGCAGAGGAGGAGCTCCTGGACATG 2234

Qy    716 GlnThrLeuThrAlaTyrIleThrTyrAlaArgGlnHisValHisProThrLeuSerAsp 735
Db   2235 GCGGTGCTAAAGGACTACATTGCCTACGCGCACAGCACCATCATGCCGCGGCTAAGTGAG 2294

Qy    736 GluAlaAlaGluAspLeuIleAsnGlyTyrValGluMetArgGlnLysGlyAsnPhePro 755
Db   2295 GAAGCCAGCCAGGCTCTCATCGAGGCTTATGTAGACATGAGGAAG-----------ATT 2342

Qy    756 GlySerSerLysLysValIleThrAlaThrProArgGlnLeuGluSerMetIleArgIle 775
Db   2343 GGCAGTAGCCGGGGAATGGTTTCTGCATACCCTCGACAGCTAGAGTCATTAATCCGCTTA 2402

Qy    776 SerGluAlaLeuAlaArgMetArgPheSerGluValValGluLysValAspAlaAlaGlu 795
Db   2403 GCAGAAGCCCATGCTAAAGTAAGATTGTCTAACAAAGTTGAAGCCATTGATGTGGAAGAG 2462

Qy    796 AlaValArgLeuLeuAspValAlaLeuGlnGlnSerAlaThrAspHisAlaThrGlyThr 815
Db   2463 GCCAAACGCCTCCATCGGGAAGCTCTGAAGCAGTCTGCAACTGATCCCCGGACTGGCATC 2522

Qy    816 IleAspMetAspLeuIleThrThrGlyValSerAlaSerGluArgIleArgArgAlaAsn 835
Db   2523 GTGGACATATCTATTCTTACTACGGGGATGAGTGCCACCTCTCGTAAACGGAAAGAAGAA 2582

Qy    836 LeuLeuAlaAlaLeuArgGluLeuIleAlaAspLysIleSerProGlySerSerSerGly 855
Db   2583 TTAGCTGAAGCATTGAAAAAGCTTATT---------TTATCTAAGGGCAAAACACCAGCT 2633

Qy    856 LeuLysThrSerGlnLeuLeuGluAspIleArgSerGlnSerSerValAspValSerLeu 875
Db   2634 CTAAAATACCAGCAACTTTTTGAAGATATTCGGGGACAATCTGACATAGCAATTACTAAA 2693

Qy    876 GlnAspIleLysAsnAlaLeuGlySerLeuGlnGlyGluGlyPheLeuThrValHisGly 895
Db   2694 GATATGTTTGAAGAAGCACTGCGTGCCCTGGCAGATGATGATTTCCTGACAGTGACTGGG 2753

Qy    896 AspIleValLys 899
Db   2754 AAGACCGTGCGC 2765

RESULT 14
AD019847
ID   AD019847 standard; cDNA; 3273 BP.
XX
AC   AD019847;
XX
DT   12-AUG-2004  (first entry)
XX
```

```
DE   Human PRO polynucleotide #386.
XX
KW   Human; PRO; gene; ss; immune related disorder;
KW   systemic lupus erythematosus; rheumatoid arthritis; osteoarthritis;
KW   juvenile chronic arthritis; systemic sclerosis; Sjogren's syndrome;
KW   vasculitis; sarcoidosis; autoimmune haemolytic anaemia;
KW   autoimmune thrombocytopenia; thyroiditis; diabetes mellitus;
KW   renal disease; demyelinating disease; central nervous system;
KW   peripheral nervous system; demyelinating polyneuropathy;
KW   Guillain-Barre syndrome;
KW   chronic inflammatory demyelinating polyneuropathy.
XX
OS   Homo sapiens.
XX
PN   WO2004043361-A2.
XX
PD   27-MAY-2004.
XX
PF   06-NOV-2003; 2003WO-US035268.
XX
PR   08-NOV-2002; 2002US-0425235P.
XX
PA   (GETH ) GENENTECH INC.
XX
PI   Fong S,  Dennis K,  Clark H,  Chiu H,  Schoenfeld J,  Williams PM;
PI   Wood WI,  Wu TD;
XX
DR   WPI; 2004-420067/39.
DR   P-PSDB; ADO19848.
XX
PT   Novel PRO polypeptide e.g., PRO69614, PRO71106, or PRO86388 useful for
PT   treating an immune related disorder such as systemic lupus erythematosus,
PT   rheumatoid arthritis, osteoarthritis, juvenile chronic arthritis or
PT   spondyloarthropathy.
XX
PS   Claim 1; SEQ ID NO 796; 1731pp; English.
XX
CC   The invention relates to human PRO polypeptides and the polynucleotides
CC   encoding them. The polypeptides and polynucleotides are useful for
CC   treating and diagnosing immune related disorders in mammals. The immune
CC   related disorders include systemic lupus erythematosus, rheumatoid
CC   arthritis, osteoarthritis, juvenile chronic arthritis, systemic
CC   sclerosis, Sjogren's syndrome, vasculitis, sarcoidosis, autoimmune
CC   haemolytic anaemia, autoimmune thrombocytopenia, thyroiditis, diabetes
CC   mellitus, immune-mediated renal disease, demyelinating diseases of the
CC   central or peripheral nervous system, demyelinating polyneuropathy,
CC   Guillain-Barre syndrome and chronic inflammatory demyelinating
CC   polyneuropathy. This sequence represents a human PRO polynucleotide of
CC   the invention.
XX
SQ   Sequence 3273 BP; 876 A; 809 C; 859 G; 729 T; 0 U; 0 Other;

Alignment Scores:
Pred. No.:              6.4e-98       Length:         3273
Score:                  1817.50       Matches:        406
Percent Similarity:     57.78%        Conservative:   151
Best Local Similarity:  42.12%        Mismatches:     256
Query Match:            39.43%        Indels:         151
DB:                     12            Gaps:           18

US-10-768-511-8 (1-901) x ADO19847 (1-3273)

Qy         38 AlaAlaSerProValAlaGlyArgArgAlaValArgGlnThrProThrSerAlaValArg 57
              |||   |||      :::||||||                  |||   :::|||   |||
Db         21 GCGACGTCGTCGGCAAGCGGCCGC---------------CTTCCACGTAACGCG---CGC 62

Qy         58 ArgArgGlyArgGluThrAspSerAlaArgArgArgArgSerArg--------------- 72
              ||||||||||||            :::||||||   |||   |||
Db         63 CGGCGGGGGAGGGCGTTGGCGCGGAGCCGACGGAACGTCCGCGCTGCGGAGCAGGGCAG 122

Qy         73 ---SerArgSerLeuGlyAsnSerValTyrSerSerProTyrAspAlaGlyThrProGly 91
                 ||||||      |||              :::   |||      ||||||||||||
Db        123 GGAAGCCGGGAGGCGGGC---CCGGCCCGAGCTTGTCCTTGTCGCGCAGGTACTCCGAGC 179

Qy         92 ThrProGlyThrProValAlaThrProVal--------------------TyrAlaThr 104
              |||      :::|||   :::||||||                         ||||||
Db        180 ACTATGTCGTCCCCGGCGTCGACCCCGAGCCGCCGCGGCAGCCGGCGTGGAAGGGCCACC 239

Qy        105 ProValGlyThrPro--------------------------------------------- 109
              |||      ||||||
Db        240 CCCGCCCAGACGCCTCGGAGTGAGGATGCCAGGTCATCTCCCTCTCAGAGACGTAGAGGC 299

Qy        109 ------------------------------------------------------------ 109

Db        300 GAGGATTCCACCTCCACGGGGGAGTTGCAGCCGATGCCAACCTCGCCTGGAGTGGACCTG 359

Qy        109 ------------------------------------------------------------ 109

Db        360 CAGAGCACTGCTGCGCAGGACGTGCTGTTTTCCAGCCCTCCCCAAATGCATTCTTCAGCT 419

Qy        110 ------------------------------------MetGlyThrProSerPheHis--- 116
                                                     ||||||||||||
Db        420 ATCCCTCTTGACTTTGATGTTAGTTCACCACTGACATACGGCACTCCCAGCTCTCGGGTA 479

Qy        117 ------------------------ArgGlyThrProGlnTyrLysGlnArgSerGluLeu 128
                                      ||||||||||||      :::||||||   :::|||
Db        480 GAGGGAACCCCAAGAAGTGGTGTTAGGGGCACACCT---GTGAGACAGAGGCCTGACCTG 536

Qy        129 GlySerGlnGlyLysProLeuHisArgArgArgArgSerGlnSerArgGluProGlyHis 148
              ||||||      |||   |||         :::|||
Db        537 GGCTCTGCACAGAAGGGCCTGCAAGTGGATCTGCAGTCT--------------------- 575

Qy        149 ArgSerProSerArgGluProSerAlaAspGlyArgProSerGluSerAlaGluProAsp 168
                                         ||||||      :::|||
Db        576 ---------------------------GACGGGGCAGCAGCAGAAGATATAGTGGCAAGT 608

Qy        169 AspThrLeuGlyGlyGluTyrAlaTyrValTrpGlyThrAsnValAsnIleProAspVal 188
              :::            |||:::         :::|||||||||:::||||||:::
Db        609 GAGCAGTCTCTAGGCCAAAAACTTGTGATCTGGGGAACAGATGTAAATGTGGCAGCATGC 668

Qy        189 LeuArgAlaIleArgArgPheLeuHisAsnTyr------------------------Arg 200
                             :::|||||||||   :::
Db        669 AAAGAAAACTTTCAGAGATTTCTTCAGCGTTTATTGACCCTCTGGCTAAAGAAGAAGAA 728

Qy        201 SerSerAlaHisAspLeuAsnSerLys---TyrIleGlnIleIleGluGluThrValGlu 219
              :::            |||:::            |||:::|||   :::   |||
Db        729 AATGTTGGCATAGATATTACTGAACCTCTATACATGCAACGACTTGGGGAGATTAATGTT 788

Qy        220 ArgGluGluAspThrLeuAsnIleAspMetSerAspIleTyrAspHisAspProAspLeu 239
                    |||      ||||||:::::            |||         |||   :::|||
Db        789 ATTGGTGAGCAATTTTTAAATGTGAACTGTGAACACATCAAATCATTTGACAAAAATTTG 848

Qy        240 TyrAlaLysIleValArgTyrProLeuAspIleIleProLeuLeuAspThrGluCysGln 259
              |||   :::::::::   ||||||   ::::::||||||         |||
Db        849 TACAGACAACTCATCTCTTACCCACAGGAAGTTATTCCAACTTTTGACATGGCTGTCAAT 908

Qy        260 GluValAlaThrSerLeuLeuPro------ThrPheGluLysHisIleGluAlaArgPro 277
              |||:::            |||               |||      |||:::   ||||||
Db        909 GAAATCTTCTTTGACCGTTACCCTGACTCAATCTTAGAACATCAGATTCAAGTAAGACCA 968

Qy        278 PheAsnLeuLysAlaSerValHisMetArgGluLeuAsnProSerAspIleAspLysLeu 297
              ||||||         :::   :::||||||   |||||||||   |||||||||:::|||
Db        969 TTCAACGCATTGAAGACTAAGAATATGAGAAACCTGAATCCAGAAGACATTGACCAGCTC 1028

Qy        298 ValSerValLysGlyMetValIleArgCysSerSerIleIleProGluIleLysGlyAla 317
              :::::::::   ||||||||||||||   |||   :::||||||||||:::::   |||
Db       1029 ATCACCATCAGCGGCATGGTGATCAGGACATCCCAGCTGATTCCCGAGATGCAGGAGGCC 1088

Qy        318 PhePheLysCysLeuValCysGlyHisSerProProLeuValThrValValLysGlyArg 337
              ||||||:::|||   ||||||   |||:::            |||   :::   :::||||||
Db       1089 TTCTTCCAGTGCCAAGTGTGTGCCCACACG---ACCCGGGTGGAGATGGACCGCGGCCGC 1145

Qy        338 ValGluGluProThrArgCysGluLysProGluCysAlaAlaArgAsnAlaMetSerLeu 357
              :::   ||||||:::   |||   :::      |||            ::::::|||:::|||
Db       1146 ATTGCAGAGCCCAGTGTGTGCGGGCGC------TGCCACACCACCCACAGCATGGCACTC 1199
```

```
Qy    358 IleHisAsnArgCysThrPheAlaAsnLysGlnIleValArgLeuGlnGluThrProAsp 377
Db   1200 ATCCACAACCGCTCCCTCTTCTCTGACAAGCAGATGATCAAGCTTCAGGAGTCTCCGGAA 1259

Qy    378 AlaIleProGluGlyGluThrProHisThrValSerMetCysLeuTyrAsnThrMetVal 397
Db   1260 GACATGCCTGCAGGGCAGACACCACACACAGTTATCCTGTTTGCTCACAATGATCTCGTT 1319

Qy    398 AspAlaValLysProGlyAspArgIleGluValThrGlyValPheLysAlaMetAlaVal 417
Db   1320 GACAAGGTCCAGCCTGGGGACAGAGTGAATGTTACAGGCATCTATCGAGCTGTGCCTATT 1379

Qy    418 ArgValGlyProAsnGlnArgThrLeuArgAlaLeuTyrLysThrTyrIleAspCysVal 437
Db   1380 CGAGTCAATCCAAGAGTGAGTAATGTGAAGTCTGTCTACAAAACCCACATTGATGTCATT 1439

Qy    438 HisValLysLysSerAspArgGlyArgLeuGlnThrGluAspProMetGluMetAspLys 457
Db   1440 CATTATCGGAAAACGGATGCAAAACGTCTGCAT-------------------------- 1472

Qy    458 GluAsnAspMetTyrAlaGlyTyrHisGluSerAspThrSerGluAlaAlaAsnGluAla 477
Db   1473 ------------------GGCCTTGATGAAGAAGCAGAACAGAAACTTTTTCAGAGAAA 1514

Qy    478 LysIleGlnLysLeuLysGluLeuSerLysLeuProGlyIleTyrAspArgLeuSerArg 497
Db   1515 CGTGTGGAATTGCTTAAGGAACTTTCCAGGAAACCAGACATTTATGAGAGGCTTGCTTCA 1574

Qy    498 SerLeuAlaProSerIleTrpGluLeuGluAspIleLysLysGlyLeuLeuCysGlnLeu 517
Db   1575 GCCTTGGCTCCAAGCATTTATGAACATGAAGATATAAAGAAGGGAATTTTGCTTCAGCTC 1634

Qy    518 PheGlyGlyLysAlaLysLysIle------ProSerGlyAlaSerPheArgGlyAspIle 535
Db   1635 TTTGGCGGGACAAGGAAGGATTTAGTCACACTGGAAGGGGCAAATTTCGGGCTGAGATC 1694

Qy    536 AsnValLeuLeuValGlyAspProGlyThrSerLysSerGlnLeuLeuGlnTyrValHis 555
Db   1695 AACATCTTGCTGTGTGGCGACCCTGGTACCAGCAAGTCCCAGCTGCTGCAGTACGTGTAC 1754

Qy    556 LysIleAlaProArgGlyIleTyrThrSerGlyArgGlySerSerAlaValGlyLeuThr 575
Db   1755 AACCTCGTCCCCAGGGGCCAGTACACGTCTGGGAAGGGCTCCAGTGCAGTTGGCCTCACT 1814

Qy    576 AlaTyrValThrLysAspProGluThrArgGluThrValLeuGluSerGlyAlaLeuVal 595
Db   1815 GCGTACGTAATGAAAGACCCTGAGACAAGGCAGCTGGTCCTGCAGACAGGTGCTCTTGTC 1874

Qy    596 LeuSerAspArgGlyIleCysCysIleAspGluPheAspLysMetSerAspAsnAlaArg 615
Db   1875 CTGAGTGACAACGGCATCTGCTGTATCGATGAGTTCGACAAGATGAATGAAAGTACAAGA 1934

Qy    616 SerMetLeuHisGluValMetGluGlnGlnThrValSerValAlaLysGlyGlyIleIle 635
Db   1935 TCGGTATTGCATGAAGTCATGGAACAGCAGACTCTGTCCATTGCAAAGGCTGGGATCATC 1994

Qy    636 AlaSerLeuAsnAlaArgThrSerValLeuAlaCysAlaAsnProSerGlySerArgTyr 655
Db   1995 TGTCAGCTCAATGCGCGCACCTCTGTCCTGGCAGCAGCAAATCCCATTGAGTCTCAGTGG 2054

Qy    656 AsnAlaArgLeuSerValIleAspAsnIleGlnLeuProProThrLeuLeuSerArgPhe 675
Db   2055 AATCCTAAAAAAACAACCATTGAAAACATCCAGCTGCCTCATACTTTATTATCAAGGTTT 2114

Qy    676 AspLeuIleTyrLeuMetLeuAspLysProAspGluGlnAsnAspArgArgLeuAlaArg 695
Db   2115 GATTTGATCTTCCTCATGCTGGACCCTCAGGACGAAGCCTATGACAGGCGTCTGGCTCAC 2174

Qy    696 HisLeuValAlaLeuHisTyrGluAsnTyrGluValSerLysGlnAspAlaLeuAspLeu 715
Db   2175 CACCTGGTCGCACTGTACTACCAGAGCGAGGAGCAGGCAGAGGAGGAGCTCCTGGACATG 2234

Qy    716 GlnThrLeuThrAlaTyrIleThrTyrAlaArgGlnHisValHisProThrLeuSerAsp 735
Db   2235 GCGGTGCTAAAGGACTACATTGCCTACGCGCACAGCACCATCATGCCGCGGCTAAGTGAG 2294

Qy    736 GluAlaAlaGluAspLeuIleAsnGlyTyrValGluMetArgGlnLysGlyAsnPhePro 755
Db   2295 GAAGCCAGCCAGGCTCTCATCGAGGCTTATGTAGACATGAGGAAG-----------ATT 2342

Qy    756 GlySerSerLysLysValIleThrAlaThrProArgGlnLeuGluSerMetIleArgIle 775
Db   2343 GGCAGTAGCCGGGGAATGGTTTCTGCATACCCTCGACAGCTAGAGTCATTAATCCGCTTA 2402

Qy    776 SerGluAlaLeuAlaArgMetArgPheSerGluValValGluLysValAspAlaAlaGlu 795
Db   2403 GCAGAAGCCCATGCTAAAGTAAGATTGTCTAACAAAGTTGAAGCCATTGATGTGGAAGAG 2462

Qy    796 AlaValArgLeuLeuAspValAlaLeuGlnGlnSerAlaThrAspHisAlaThrGlyThr 815
Db   2463 GCCAAACGCCTCCATCGGGAAGCTCTGAAGCAGTCTGCAACTGATCCCCGGACTGGCATC 2522

Qy    816 IleAspMetAspLeuIleThrThrGlyValSerAlaSerGluArgIleArgArgAlaAsn 835
Db   2523 GTGGACATATCTATTCTTACTACGGGGATGAGTGCCACCTCTCGTAAACGGAAAGAAGAA 2582

Qy    836 LeuLeuAlaAlaLeuArgGluLeuIleAlaAspLysIleSerProGlySerSerSerGly 855
Db   2583 TTAGCTGAAGCATTGAAAAAGCTTATT---------TTATCTAAGGGCAAAACACCAGCT 2633

Qy    856 LeuLysThrSerGlnLeuLeuGluAspIleArgSerGlnSerSerValAspValSerLeu 875
Db   2634 CTAAAATACCAGCAACTTTTGAAGATATTCGGGGACAATCTGACATAGCAATTACTAAA 2693

Qy    876 GlnAspIleLysAsnAlaLeuGlySerLeuGlnGlyGluGlyPheLeuThrValHisGly 895
Db   2694 GATATGTTTGAAGAAGCACTGCGTGCCCTGGCAGATGATGATTTCCTGACAGTGACTGGG 2753

Qy    896 AspIleValLys 899
Db   2754 AAGACCGTGCGC 2765
```

```
RESULT 15
ADR26059
ID   ADR26059 standard; DNA; 3273 BP.
XX
AC   ADR26059;
XX
DT   21-OCT-2004  (first entry)
XX
DE   Breast cancer prognosis marker #1920.
XX
KW   ds; breast cancer; prognosis; gene expression; diagnosis.
XX
OS   Homo sapiens.
XX
PN   WO2004065545-A2.
XX
PD   05-AUG-2004.
XX
PF   15-JAN-2004; 2004WO-US001100.
XX
PR   15-JAN-2003; 2003US-00342887.
XX
PA   (ROSE-) ROSETTA INPHARMATICS LLC.
PA   (NECA-) NETHERLANDS CANCER INST.
XX
PI   Van't Veer LJ,  He Y;
XX
DR   WPI; 2004-593473/57.
XX
PT   Classifying a breast cancer patient according to prognosis comprises
PT   determining the similarity between the level of expression of each of
PT   five genes in a cell sample taken from patient, to control levels.
XX
PS   Disclosure; SEQ ID NO 1920; 226pp; English.
```

```
XX
CC   The invention relates to a method of classifying a breast cancer patient
CC   according to prognosis by determining the similarity between the level of
CC   expression of each of five genes for which markers are listed in the
CC   specification, in a cell sample taken from the breast cancer patient, to
CC   control levels of expression for each respective five genes to obtain a
CC   patient similarity value. The methods are useful for classifying a breast
CC   cancer patient according to prognosis. Kits and computer program products
CC   are useful for data analysis using the diagnostic, prognostic and
CC   statistical methods of the invention. This sequence corresponds to a
CC   marker used in the method of the invention.
XX
SQ   Sequence 3273 BP; 876 A; 809 C; 859 G; 729 T; 0 U; 0 Other;

Alignment Scores:
Pred. No.:              6.4e-98        Length:        3273
Score:                  1817.50        Matches:       406
Percent Similarity:     57.78%         Conservative:  151
Best Local Similarity:  42.12%         Mismatches:    256
Query Match:            39.43%         Indels:        151
DB:                     13             Gaps:          18

US-10-768-511-8 (1-901) x ADR26059 (1-3273)

Qy        38 AlaAlaSerProValAlaGlyArgArgAlaValArgGlnThrProThrSerAlaValArg 57
Db        21 GCGACGTCGTCGGCAAGCGGCCGC---------------CTTCCACGTAACGCG---CGC 62

Qy        58 ArgArgGlyArgGluThrAspSerAlaArgArgArgArgSerArg--------------- 72
Db        63 CGGCGGGGGAGGGCGTTGGCGCGGAGCCGACGGGAACGTCCGCGCTGCGGAGCAGGGCAG 122

Qy        73 ---SerArgSerLeuGlyAsnSerValTyrSerSerProTyrAspAlaGlyThrProGly 91
Db       123 GGAAGCCGGGAGGCGGGC---CCGGCCCGAGCTTGTCCTTGTCGCGCAGGTACTCCGAGC 179

Qy        92 ThrProGlyThrProValAlaThrProVal--------------------TyrAlaThr 104
Db       180 ACTATGTCGTCCCCGGCGTCGACCCCGAGCCGCCGCGGCAGCCGGCGTGGAAGGGCCACC 239

Qy       105 ProValGlyThrPro--------------------------------------------- 109
Db       240 CCCGCCCAGACGCCTCGGAGTGAGGATGCCAGGTCATCTCCCTCTCAGAGACGTAGAGGC 299

Qy       109 ------------------------------------------------------------ 109
Db       300 GAGGATTCCACCTCCACGGGGGAGTTGCAGCCGATGCCAACCTCGCCTGGAGTGGACCTG 359

Qy       109 ------------------------------------------------------------ 109
Db       360 CAGAGCACTGCTGCGCAGGACGTGCTGTTTTCCAGCCCTCCCCAAATGCATTCTTCAGCT 419

Qy       110 ------------------------------------MetGlyThrProSerPheHis--- 116
Db       420 ATCCCTCTTGACTTTGATGTTAGTTCACCACTGACATACGGCACTCCCAGCTCTCGGGTA 479

Qy       117 -----------------------ArgGlyThrProGlnTyrLysGlnArgSerGluLeu 128
Db       480 GAGGGAACCCCAAGAAGTGGTGTTAGGGGCACACCT---GTGAGACAGAGGCCTGACCTG 536

Qy       129 GlySerGlnGlyLysProLeuHisArgArgArgArgSerGlnSerArgGluProGlyHis 148
Db       537 GGCTCTGCACAGAAGGGCCTGCAAGTGGATCTGCAGTCT-------------------- 575

Qy       149 ArgSerProSerArgGluProSerAlaAspGlyArgProSerGluSerAlaGluProAsp 168
Db       576 --------------------------GACGGGGCAGCAGCAGAAGATATAGTGGCAAGT 608

Qy       169 AspThrLeuGlyGlyGluTyrAlaTyrValTrpGlyThrAsnValAsnIleProAspVal 188
Db       609 GAGCAGTCTCTAGGCCAAAAACTTGTGATCTGGGGAACAGATGTAAATGTGGCAGCATGC 668

Qy       189 LeuArgAlaIleArgArgPheLeuHisAsnTyr-----------------------Arg 200
Db       669 AAAGAAAACTTTCAGAGATTTCTTCAGCGTTTTATTGACCCTCTGGCTAAAGAAGAAGAA 728

Qy       201 SerSerAlaHisAspLeuAsnSerLys---TyrIleGlnIleIleGluGluThrValGlu 219
Db       729 AATGTTGGCATAGATATTACTGAACCTCTATACATGCAACGACTTGGGGAGATTAATGTT 788

Qy       220 ArgGluGluAspThrLeuAsnIleAspMetSerAspIleTyrAspHisAspProAspLeu 239
Db       789 ATTGGTGAGCAATTTTTAAATGTGAACTGTGAACACATCAAATCATTTGACAAAAATTTG 848

Qy       240 TyrAlaLysIleValArgTyrProLeuAspIleIleProLeuLeuAspThrGluCysGln 259
Db       849 TACAGACAACTCATCTCTTACCCACAGGAAGTTATTCCAACTTTTGACATGGCTGTCAAT 908

Qy       260 GluValAlaThrSerLeuLeuPro------ThrPheGluLysHisIleGluAlaArgPro 277
Db       909 GAAATCTTCTTTGACCGTTACCCTGACTCAATCTTAGAACATCAGATTCAAGTAAGACCA 968

Qy       278 PheAsnLeuLysAlaSerValHisMetArgGluLeuAsnProSerAspIleAspLysLeu 297
Db       969 TTCAACGCATTGAAGACTAAGAATATGAGAAACCTGAATCCAGAAGACATTGACCAGCTC 1028

Qy       298 ValSerValLysGlyMetValIleArgCysSerSerIleIleProGluIleLysGlyAla 317
Db      1029 ATCACCATCAGCGGCATGGTGATCAGGACATCCCAGCTGATTCCCGAGATGCAGGAGGCC 1088

Qy       318 PhePheLysCysLeuValCysGlyHisSerProProLeuValThrValValLysGlyArg 337
Db      1089 TTCTTCCAGTGCCAAGTGTGTGCCCACACG---ACCCGGGTGGAGATGGACCGCGGCCGC 1145

Qy       338 ValGluGluProThrArgCysGluLysProGluCysAlaAlaArgAsnAlaMetSerLeu 357
Db      1146 ATTGCAGAGCCCAGTGTGTGCGGGCGC------TGCCACACCACCCACAGCATGGCACTC 1199

Qy       358 IleHisAsnArgCysThrPheAlaAsnLysGlnIleValArgLeuGlnGluThrProAsp 377
Db      1200 ATCCACAACCGCTCCCTCTTCTCTGACAAGCAGATGATCAAGCTTCAGGAGTCTCCGGAA 1259

Qy       378 AlaIleProGluGlyGluThrProHisThrValSerMetCysLeuTyrAsnThrMetVal 397
Db      1260 GACATGCCTGCAGGGCAGACACCACACACAGTTATCCTGTTTGCTCACAATGATCTCGTT 1319

Qy       398 AspAlaValLysProGlyAspArgIleGluValThrGlyValPheLysAlaMetAlaVal 417
Db      1320 GACAAGGTCCAGCCTGGGGACAGAGTGAATGTTACAGGCATCTATCGAGCTGTGCCTATT 1379

Qy       418 ArgValGlyProAsnGlnArgThrLeuArgAlaLeuTyrLysThrTyrIleAspCysVal 437
Db      1380 CGAGTCAATCCAAGAGTGAGTAATGTGAAGTCTGTCTACAAAACCCACATTGATGTCATT 1439

Qy       438 HisValLysLysSerAspArgGlyArgLeuGlnThrGluAspProMetGluMetAspLys 457
Db      1440 CATTATCGGAAAACGGATGCAAAACGTCTGCAT-------------------------- 1472

Qy       458 GluAsnAspMetTyrAlaGlyTyrHisGluSerAspThrSerGluAlaAlaAsnGluAla 477
Db      1473 -----------------GGCCTTGATGAAGAAGCAGAACAGAAACTTTTTCAGAGAAA 1514

Qy       478 LysIleGlnLysLeuLysGluLeuSerLysLeuProGlyIleTyrAspArgLeuSerArg 497
Db      1515 CGTGTGGAATTGCTTAAGGAACTTTCCAGGAAACCAGACATTTATGAGAGGCTTGCTTCA 1574

Qy       498 SerLeuAlaProSerIleTrpGluLeuGluAspIleLysLysGlyLeuLeuCysGlnLeu 517
Db      1575 GCCTTGGCTCCAAGCATTTATGAACATGAAGATATAAAGAAGGGAATTTTGCTTCAGCTC 1634

Qy       518 PheGlyGlyLysAlaLysLysIle------ProSerGlyAlaSerPheArgGlyAspIle 535
Db      1635 TTTGGCGGGACAAGGAAGGATTTAGTCACACTGGAAGGGGCAAATTTCGGGCTGAGATC 1694

Qy       536 AsnValLeuLeuValGlyAspProGlyThrSerLysSerGlnLeuLeuGlnTyrValHis 555
```

```
Db      1695 AACATCTTGCTGTGTGGCGACCCTGGTACCAGCAAGTCCCAGCTGCTGCAGTACGTGTAC 1754

Qy       556 LysIleAlaProArgGlyIleTyrThrSerGlyArgGlySerSerAlaValGlyLeuThr 575
                 :::  |||||||||  |||||||||||:::|||||||||||||||||||||||
Db      1755 AACCTCGTCCCCAGGGGCCAGTACACGTCTGGGAAGGGCTCCAGTGCAGTTGGCCTCACT 1814

Qy       576 AlaTyrValThrLysAspProGluThrArgGluThrValLeuGluSerGlyAlaLeuVal 595
             |||||||||  ||||||||||||||||:::  ||||||::::::|||||||||||
Db      1815 GCGTACGTAATGAAAGACCCTGAGACAAGGCAGCTGGTCCTGCAGACAGGTGCTCTTGTC 1874

Qy       596 LeuSerAspArgGlyIleCysCysIleAspGluPheAspLysMetSerAspAsnAlaArg 615
             |||||||||  ||||||||||||||||||||||||||||||:::::::::   |||
Db      1875 CTGAGTGACAACGGCATCTGCTGTATCGATGAGTTCGACAAGATGAATGAAAGTACAAGA 1934

Qy       616 SerMetLeuHisGluValMetGluGlnGlnThrValSerValAlaLysGlyGlyIleIle 635
             |||:::||||||||||||||||||||||||||||:::|||:::||||||  |||||||||
Db      1935 TCGGTATTGCATGAAGTCATGGAACAGCAGACTCTGTCCATTGCAAAGGCTGGGATCATC 1994

Qy       636 AlaSerLeuAsnAlaArgThrSerValLeuAlaCysAlaAsnProSerGlySerArgTyr 655
                   ||||||||||||||||||||||||||  |||||||||    |||::::::
Db      1995 TGTCAGCTCAATGCGCGCACCTCTGTCCTGGCAGCAGCAAATCCCATTGAGTCTCAGTGG 2054

Qy       656 AsnAlaArgLeuSerValIleAspAsnIleGlnLeuProProThrLeuLeuSerArgPhe 675
             |||   :::   :::   |||:::||||||||||||||||  ||||||||||||||||||
Db      2055 AATCCTAAAAAAACAACCATTGAAAACATCCAGCTGCCTCATACTTTATTATCAAGGTTT 2114

Qy       676 AspLeuIleTyrLeuMetLeuAspLysProAspGluGlnAsnAspArgArgLeuAlaArg 695
             |||||||||:::||||||||||||        ||||||      ||||||||||||||
Db      2115 GATTTGATCTTCCTCATGCTGGACCCTCAGGACGAAGCCTATGACAGGCGTCTGGCTCAC 2174

Qy       696 HisLeuValAlaLeuHisTyrGluAsnTyrGluValSerLysGlnAspAlaLeuAspLeu 715
             |||||||||||||||:::|||::::::   |||   ::::::::::::   ||||||:::
Db      2175 CACCTGGTCGCACTGTACTACCAGAGCGAGGAGCAGGCAGAGGAGGAGCTCCTGGACATG 2234

Qy       716 GlnThrLeuThrAlaTyrIleThrTyrAlaArgGlnHisValHisProThrLeuSerAsp 735
                   |||     ||||||   ||||||            :::   |||   ||||||:::
Db      2235 GCGGTGCTAAAGGACTACATTGCCTACGCGCACAGCACCATCATGCCGCGGCTAAGTGAG 2294

Qy       736 GluAlaAlaGluAspLeuIleAsnGlyTyrValGluMetArgGlnLysGlyAsnPhePro 755
             ||||||::::::   ||||||      ||||||:::||||||:::
Db      2295 GAAGCCAGCCAGGCTCTCATCGAGGCTTATGTAGACATGAGGAAG------------ATT 2342

Qy       756 GlySerSerLysLysValIleThrAlaThrProArgGlnLeuGluSerMetIleArgIle 775
             |||||||||:::   :::::::::|||   ||||||||||||||||||:::|||||||:::
Db      2343 GGCAGTAGCCGGGGAATGGTTTCTGCATACCCTCGACAGCTAGAGTCATTAATCCGCTTA 2402

Qy       776 SerGluAlaLeuAlaArgMetArgPheSerGluValValGluLysValAspAlaAlaGlu 795
             :::||||||   |||::::::|||   |||         ||||||   :::|||   |||
Db      2403 GCAGAAGCCCATGCTAAAGTAAGATTGTCTAACAAAGTTGAAGCCATTGATGTGGAAGAG 2462

Qy       796 AlaValArgLeuLeuAspValAlaLeuGlnGlnSerAlaThrAspHisAlaThrGlyThr 815
             |||   ||||||            ||||||:::||||||||||||||||      ||||||
Db      2463 GCCAAACGCCTCCATCGGGAAGCTCTGAAGCAGTCTGCAACTGATCCCCGGACTGGCATC 2522

Qy       816 IleAspMetAspLeuIleThrThrGlyValSerAlaSerGluArgIleArgArgAlaAsn 835
             :::|||:::   ::::::|||||||||:::||||||:::   |||   |||:::
Db      2523 GTGGACATATCTATTCTTACTACGGGGATGAGTGCCACCTCTCGTAAACGGAAAGAAGAA 2582

Qy       836 LeuLeuAlaAlaLeuArgGluLeuIleAlaAspLysIleSerProGlySerSerSerGly 855
             |||      ||||||::::::||||||         :::|||   |||   :::
Db      2583 TTAGCTGAAGCATTGAAAAAGCTTATT---------TTATCTAAGGGCAAAACACCAGCT 2633

Qy       856 LeuLysThrSerGlnLeuLeuGluAspIleArgSerGlnSerSerValAspValSerLeu 875
             ||||||      ||||||   ||||||||||||   ||||||   :::   ::::::
Db      2634 CTAAAATACCAGCAACTTTTTGAAGATATTCGGGGACAATCTGACATAGCAATTACTAAA 2693

Qy       876 GlnAspIleLysAsnAlaLeuGlySerLeuGlnGlyGluGlyPheLeuThrValHisGly 895
                   :::   ||||||   :::|||      :::   ||||||||||||      |||
Db      2694 GATATGTTTGAAGAAGCACTGCGTGCCCTGGCAGATGATGATTTCCTGACAGTGACTGGG 2753

Qy       896 AspIleValLys 899
                   |||:::
Db      2754 AAGACCGTGCGC 2765
```

Search completed: December 6, 2005, 10:15:30
Job time : 972 secs

```
                   :||| : :||||  | :| |  |      | :||::|   | : ||:   | |
Db        257 RGENTRLAQPGDHVGITGVFLPM-LRTGFRQLVQGLLSETYLECHRLVKMNKSEDDELGT 315

Qy        450 EDPMEMDKENDMYAGYHESDTSEAANEAKIQKLKELSKLPGIYDRLSRSLAPSIWELEDI 509
              |:  |                        ::|:::::   |::|: |:|| |:  ||:
Db        316 EELSE------------------------EELRQITE-EDFYEKLAASIAPEIYGHEDV 349

Qy        510 KKGLLCQLFGGKAKKIPSGASFRGDINVLLVGDPGTSKSQLLQYVHKIAPRGIYTSGRGS 569
              || || | ||          | |     ||:||: |:|||| :||||| |: ::||  ||:||||
Db        350 KKALLLLLVGG-VDNSPRGMKIRGNINICLMGDPGVAKSQLLSYIDRLAPRSQYTTGRGS 408

Qy        570 SAVGLTAYVTKDPETRETVLESGALVLSDRGICCIDEFDKMSDNARSMLHEVMEQQTVSV 629
              | ||||| | ||| | |  || |||||:|:|:||||||||| |   |: :||||||||:|:
Db        409 SGVGLTAAVMKDPVTGEMTLEGGALVLADQGVCCIDEFDKMMDTDRTAIHEVMEQQTISI 468

Qy        630 AKGGIIASLNARTSVLACANPSGSRYNARLSVIDNIQLPPTLLSRFDLIYLMLDKPDEQN 689
              || ||: :|||| |:|| |||:  ||| : :|  |||||  ||||||:::|: ||||  |
Db        469 AKAGIMTTLNARCSILAAANPAYGRYNPKKTVEQNIQLPAALLSRFDVLWLIQDKPDRDN 528

Qy        690 DRRLARHLVALHYENYEVSKQ-DALDLQTLTAYITYARQHVHPTLSDEAAEDLINGYVEM 748
              | |||:|:  :|  : :   |    ||:: :  |||  ::    | : :   |: |   ||||
Db        529 DLRLAQHITYVHQHSKQPPSQFQPLDMKLMRRYITMCKRK-QPAIPEALADYLTAAYVEM 587

Qy        749 RQKGNFPGSSKKVITATPRQLESMIRISEALARMRFSEVVEKVDAAEAVRLLDVALQQSA 808
              |::    ::| :    :  | |  |::|:| ||||:  :|||||  |  ||:||::::
Db        588 RKEAR---TNKDMTFTSARTLLSVLRLSTALARLRLEDVVEKEDVNEAMRLMEMSKDSLL 644

Qy        809 TDHATGTIDMDLITTGVSASERIRRANLL-AALRELIADK 847
               |                  |    | |:::: : :||:: :|
Db        645 GDKG------------HTSRTQRPADVIFSTIREMVPEK 671
```

RESULT 15
JC4580
replication licensing factor MCM7 - mouse
N;Alternate names: cell division control protein CDC47
C;Species: Mus musculus (house mouse)
C;Date: 13-Aug-1999 #sequence_revision 13-Aug-1999 #text_change 09-Jul-2004
C;Accession: JC4580
R;Takizawa, N.; Kimura, H.; Sugimoto, K.
Gene 167, 343-344, 1995
A;Title: Sequence of mouse CDC47, a member of the minichromosome maintenance (Mcm) famil
A;Reference number: JC4580; MUID:96144303; PMID:8566808
A;Accession: JC4580
A;Molecule type: mRNA
A;Residues: 1-719 <TAK>
A;Cross-references: UNIPROT:Q61881; UNIPARC:UPI0000029103; DDBJ:D26091; NID:g1136746; PI
C;Comment: The complex of six MCM proteins is one of several proteins that must be bound
 phosphorylated and dissociate from the chromatin.
C;Genetics:
A;Gene: cdc47
C;Complex: The predominant form is a heterohexamer of MCM2, MCM3, MCM4, MCM5, MCM6, and
C;Function:
A;Description: part of the replication licensing system that permits DNA replication to
C;Superfamily: replication licensing factor MCM7; MCM homology
C;Keywords: cell cycle control; DNA replication initiation; heterohexamer; nucleus; phos
F;149-636/Domain: MCM homology <MCM>

```
  Query Match             21.0%;  Score 966;  DB 1;  Length 719;
  Best Local Similarity   32.6%;  Pred. No. 1.6e-46;
  Matches  236;  Conservative  132;  Mismatches  221;  Indels  134;  Gaps   19;

Qy        192 IRRFLHN--YRSSAHDLNSKYIQIIEETVEREEDTLNIDMSDIYDHDPDLYAKIVRYPLD 249
              :::||      | :         ||   :      ||:  | :|: || : ||:
Db         13 VKKFLQEFYYENELGKKQFKYGTQLVHLAHREQVALYVDLDDIAEDDPE---------- 61

Qy        250 IIPLLDTECQ----------EVATSLLPTFEK-----------HIEAR----------- 276
                 |:|: |:           :|   ||| :::            :|| |
Db         62 ---LVDSICENAKRYSRLFGDVVQELLPEYKEKEVVNKDVLDVYIEHRLMMEQRSRDPGA 118

Qy        277 ----------------------PFNLKASVHMRELNPSDIDKLVSVKGMVIRCSSIIPEI 314
                                    | : |  | :||:    : ||::|:|:| | | : | :
Db        119 VRNPQNQYPSELMRRFELYFRGPSSSKPRV-IREVRADSVGKLLTVRGIVTRVSEVKPRM 177

Qy        315 KGAFFKCLVCG---HSPPLVTVVKGRVEEPT-----RCEKPECAARNA---MSLIHNRCT 363
               | : |  ||   : |         :: ||          |   ||    :    : |
Db        178 VVATYTCDQCGAETYQP---------IQSPTFMPLIMCPSQECQTNRSGGRLYLQTRGSK 228

Qy        364 FANKQIVRLQETPDAIPEGETPHTVSMCLYNTMVDAVKPGDRIEVTGVFKAMAVRVGPNQ 423
              |    | :::||  | :|  |  | :::: |          :||| : |||:|  : :| |  |
Db        229 FVKFQEMKIQEHSDQVPVGNIPRSITVVLEGENTRIAQPGDHVSVTGIFLPV-LRTGFQQ 287

Qy        424 RTLRALYKTYIDCVHVKKSDRGRLQTEDPMEMDKENDMYAGYHESDTSEAANEAKIQKLK 483
                      | :||::   : |           | | :|  :|         | |   ::||
Db        288 MAQGLLSETYLEAHWIVK------------MTKSDDDVSG---------AGELSSEELK 325

Qy        484 ELSKLPGIYDRLSRSLAPSIWELEDIKKGLLCQLFGGKAKKIPSGASFRGDINVLLVGDP 543
              ::::      |::|: |:|| |:  ||:|| ||  | ||  :  |    ||:|:: |:|||
Db        326 QIAE-EDFYEKLAASIAPEIYGHEDVKKALLLLLVGG-VDQSPQGMKIRGNIHICLMGDP 383

Qy        544 GTSKSQLLQYVHKIAPRGIYTSGRGSSAVGLTAYVTKDPETRETVLESGALVLSDRGICC 603
              | :||||| |: ::|||  ||:|||||  |||||  | :|   :   || |||||:|:|:||
Db        384 GVAKSQLLSYIDRLAPRSQYTTGRGSSGVGLTAAVLRDSVSGELTLEGGALVLADQGVCC 443

Qy        604 IDEFDKMSDNARSMLHEVMEQQTVSVAKGGIIASLNARTSVLACANPSGSRYNARLSVID 663
              |||||||::   |: :|||||||:|:|| ||: :|||| |:|| |||:  |||  | |:
Db        444 IDEFDKMAEADRTAIHEVMEQQTISIAKAGILTTLNARCSILAAANPAYGRYNPRRSLEQ 503

Qy        664 NIQLPPTLLSRFDLIYLMLDKPDEQNDRRLARHLVALHYENYEVSKQ-DALDLQTLTAYI 722
              |:|||  |||||||::|: |:|| || |||:|:   :|   : :     | : ||:: :  ||
Db        504 NVQLPAALLSRFDLLWLIQDRPDRDNDLRLAQHITYVHQHSRQPPAQFEPLDMKLMRRYI 563

Qy        723 TYARQHVHPTLSDEAAEDLINGYVEMRQKGNFPGSSKKVITATPRQLESMIRISEALARM 782
                :   ||: :  |: :   ||||::        :||        : | | :::|:| ||||:
Db        564 AMCHER-QPTVPESLADYITAAYVEMRREAR---ASKDATYTSARTLLAILRLSTALARL 619

Qy        783 RFSEVVEKVDAAEAVRLLDVALQQSATDHATGTIDMDLITTGVSASERIRRANLLAALRE 842
               | ::||| |   ||:||::::                |   |  :|    :    : | :||
Db        620 RMVDIVEKEDVNEAIRLMEMSK-----------DSLLGEKGQTARTQRPADVIFATIRE 667

Qy        843 LIA 845
              |::
Db        668 LVS 670

Search completed: December  7, 2005, 17:43:43
Job time : 49 secs
```

```
Db        425 YTTGRGSSGVGLTAAVMRDPVTDEMVLEGGALVLADNGICCIDEFDKMDESDRTAIHEVM 484

Qy        623 EQQTVSVAKGGIIASLNARTSVLACANPSGSRYNARLSVIDNIQLPPTLLSRFDLIYLML 682

Db        485 EQQTISISKAGITTTLNARTSILAAANPLYGRYNPKVAPIHNINLPAALLSRFDILFLIL 544

Qy        683 DKPDEQNDRRLARHLVALHYENYEVSKQD--ALDLQTLTAYITYARQHVHPTLSDEAAED 740

Db        545 DTPSRETDEHLAQHVTYVHMHN-EQPKMDFEPLDPNMIRHYISSARQY-RPVVPKDVCDY 602

Qy        741 LINGYVEMRQ-KGNFPGSSKKVITATPRQLESMIRISEALARMRFSEVVEKVDAAEAVRL 799

Db        603 VTGAYVQLRQNQKRDEANERQFAHTTPRTLLAILRMGQALARLRFSNRVEIGDVDEALRL 662

Qy        800 LDVA-------LQQSATDHATGT----IDMDLITTGVSASERIRRANLLAALRELIADKI 848

Db        663 MSVSKSSLYDDLDPSSHDTTITSKIYKIIRDMLNS-IPDVEGNERSLTLRAIRERVLAK- 720

Qy        849 SPGSSSGLKTSQLLEDIRSQSSVDVSL 875

Db        721 ------GFTEDHLINTIQEYTDLGVLL 741
```

RESULT 13
B72775
probable DNA replication licensing factor mcm APE0188 - Aeropyrum pernix (strain K1)
C;Species: Aeropyrum pernix
C;Date: 20-Aug-1999 #sequence_revision 20-Aug-1999 #text_change 09-Jul-2004
C;Accession: B72775
R;Kawarabayasi, Y.; Hino, Y.; Horikawa, H.; Yamazaki, S.; Haikawa, Y.; Jin-no, K.; Takah
awa, H.; Takamiya, M.; Masuda, S.; Funahashi, T.; Tanaka, T.; Kudoh, Y.; Yamazaki, J.; K
DNA Res. 6, 83-101, 1999
A;Title: Complete genome sequence of an aerobic hyper-thermophilic Crenarchaeon, Aeropyr
A;Reference number: A72450; MUID:99310339; PMID:10382966
A;Accession: B72775
A;Status: preliminary
A;Molecule type: DNA
A;Residues: 1-699 <KAW>
A;Cross-references: UNIPROT:Q9YFR1; UNIPARC:UPI000005DA3C; DDBJ:AP000058; NID:g5103388;
A;Experimental source: strain K1
C;Genetics:
A;Gene: APE0188
C;Superfamily: replication licensing factor MCM7; MCM homology

```
  Query Match            21.1%;  Score 973;  DB 2;  Length 699;
  Best Local Similarity  32.2%;  Pred. No. 6.2e-47;
  Matches  237;  Conservative  153;  Mismatches  281;  Indels   66;  Gaps   17;

Qy        168 DDTLGGEYAYVWGTNVNIPDVLRAIRRFLHNYRSSAHDLNSKYIQIIEETVEREEDTLNI 227

Db          5 EEMLSGEETLAVG---------ERFKTFLENFRTEEGKL--KYVEAIRRMINYEETSLEV 53

Qy        228 DMSDIYDHDPDLYAKIVRYPLDIIPLLDTECQEVATSLLPTFEKHIEARPF-----NLKA 282

Db         54 EFKDLYRYDPLLSEILLEKPREFLKEASEALKEIVAQESP---EYAQGRVFTPRPTGLFD 110

Qy        283 SVHMRELNPSDIDKLVSVKGMVIRCSSIIPEIKGAFFKCLVCGHS---PPLVTVVKG-RV 338

Db        111 TERIRDIGSDHVGKLVQINGIVTRMHPRATRMVRARFRHDRCGAEFWWPANEDEVLGERI 170

Qy        339 EEPTRCEKPECA-ARNAMSLIHNRCTFANKQIVRLQETPDAIPEGETPHTVSMCLYNTMV 397

Db        171 ERPSIC--PVCGEGGGKFTLVRDKSLYIDWQKIMVQERPEDVPGGQIPRSIEVHLSRDLV 228

Qy        398 DAVKPGDRIEVTGVFKAMAVRVG-PNQRTLRALYKTYIDCVHVKKSDRGRLQTEDPMEMD 456

Db        229 EKVRPGDRVKIVGV-------VGLQSFSSSSTLYSLYMEANSILLEEK----ILEEVSIT 277

Qy        457 KENDMYAGYHESDTSEAANEAKIQKLKELSKLPGIYDRLSRSLAPSIWELEDIKKGLLCQ 516

Db        278 REDE------------------EKILQLSRDPWIKEKIIASIAPTIYGHWDLKEAIALL 318

Qy        517 LFGGKAKKIPSGASFRGDINVLLVGDPGTSKSQLLQYVHKIAPRGIYTSGRGSSAVGLTA 576

Db        319 LFGGVPKQRPDGTRTRGDIHVLFVGDPGVAKSQLLQSTAQVAPRVVYTTGKGSTAAGLTA 378

Qy        577 YVTKDPETRETVLESGALVLSDRGICCIDEFDKMSDNARSMLHEVMEQQTVSVAKGGIIA 636

Db        379 AVLRDPRTGEYFLEAGALVLADGGIAVIDEFDKMSKEDRGVIHEAMEQQTVSIAKAGIKA 438

Qy        637 SLNARTSVLACANPSGSRYNARLSVIDNIQLPPTLLSRFDLIYLMLDKPDEQNDRRLARH 696

Db        439 TLSARASLLAAGNPKFGYYDPSRSFVDNVDLPAPIISRFDLIFVVRDVIERSRDEMLASY 498

Qy        697 LVALHYENYEVSKQDALDLQTLTAYITYARQHVHPTLSDEAAEDLINGYVEMRQKGNFPG 756

Db        499 VLETH-TNVELFKPE-IDPDLLRKYIAFARKHVKPRLTPQAKKLLKDFYVEMRSSALHHS 556

Qy        757 SSK--KVITATPRQLESMIRISEALARMRFSEVVEKVDAAEAVRLLDVALQQSATDHATG 814

Db        557 SQEGAKPVPITTRQLEALIRLTEAHARMSLKQEATEEDAIAAIRIMTSVLQSIGLDLETG 616

Qy        815 TIDMDLITTGVSASERIRRANLLAALRELIADKISPGSSSGLKTSQLLEDIRSQSSVDVS 874

Db        617 EIDIGIIMTGASFRSRKIMSEVLDLIKSIVEEE--RGGQGCVRASEI---VRRLGEKNIP 671

Qy        875 LQDIKNALGSLQGEGFL 891

Db        672 EEKVRDAIDKLYRQGLI 688
```

RESULT 14
T47221
replication licensing factor MCM7 [validated] - African clawed frog
N;Alternate names: replication licensing factor CDC47
C;Species: Xenopus laevis (African clawed frog)
C;Date: 20-Apr-2000 #sequence_revision 20-Apr-2000 #text_change 09-Jul-2004
C;Accession: T47221
R;Kubota, Y.; Mimura, S.; Nishimoto, S.; Masuda, T.; Nojima, H.; Takisawa, H.
EMBO J. 16, 3320-3331, 1997
A;Title: Licensing of DNA replication by a multi-protein complex of MCM/P1 proteins in
A;Reference number: Z24400; MUID:97357318; PMID:9214647
A;Accession: T47221
A;Status: preliminary; translated from GB/EMBL/DDBJ
A;Molecule type: mRNA
A;Residues: 1-720 <KUB>
A;Cross-references: UNIPROT:O42591; UNIPARC:UPI00000FC326; EMBL:U44051; NID:g2231176; P
C;Complex: the replication licensing complex consists of at least six proteins; MCM2 (P
alidated, MUID:97357318]
C;Function:
A;Description: MCM7 is a component of the replication licensing factor that permits DNA
C;Superfamily: replication licensing factor MCM7; MCM homology
F;148-635/Domain: MCM homology <MCM>

```
  Query Match            21.0%;  Score 969;  DB 2;  Length 720;
  Best Local Similarity  34.3%;  Pred. No. 1.1e-46;
  Matches  240;  Conservative  119;  Mismatches  221;  Indels  120;  Gaps   18;

Qy        209 KYIQIIEETVEREEDTLNIDMSDIYDHDPDLYAKIVRYPLDIIPLLDTECQEVATSLLPT 268

Db         31 KYGVQLANIAHREQVALYIDLDDLAEEDPELVDAICENTRRYTNLFADAVQE----LLPQ 86

Qy        269 FEK-----------HIEAR----------------------------------PFNLKAS 283

Db         87 YKEREVVHKDALDVYIEHRLMMEQRGRDPNEMRDSQNQYPPELMRRFELYFKAPSSSKAR 146

Qy        284 VHMRELNPSDIDKLVSVKGMVIRCSSIIPEIKGAFFKCLVCG---HSPPLVTVVKGRVEE 340

Db        147 V-VRDVKADSIGKLVNIRGIVTRVTEVKPMMVVATYTCDQCGAETYQP---------IQS 196

Qy        341 PT-----RCEKPECAARNA---MSLIHNRCTFANKQIVRLQETPDAIPEGETPHTVSMCL 392

Db        197 PTFMPLIMCPSRECQTNRSGGRLYLQTRGSKFIKFQELKIQEHSDQVPVWNIPRCMSVYV 256

Qy        393 YNTMVDAVKPGDRIEVTGVFKAMAVRVGPNQRTLRALYKTYIDC---VHVKKSDRGRLQT 449
```

```
C;Accession: D69103
R;Smith, D.R.; Doucette-Stamm, L.A.; Deloughery, C.; Lee, H.; Dubois, J.; Aldredge, T.;
; Qiu, D.; Spadafora, R.; Vicaire, R.; Wang, Y.; Wierzbowski, J.; Gibson, R.; Jiwani, N.
ki, S.; Church, G.M.; Daniels, C.J.; Mao, J.; Rice, P.; Noelling, J.; Reeve, J.N.
J. Bacteriol. 179, 7135-7155, 1997
A;Title: Complete genome sequence of Methanobacterium thermoautotrophicum Delta H: funct
A;Reference number: A69000; MUID:98037514; PMID:9371463
A;Accession: D69103
A;Status: nucleic acid sequence not shown; translation not shown
A;Molecule type: DNA
A;Residues: 1-666 <MTH>
A;Cross-references: UNIPROT:O27798; UNIPARC:UPI0000066785; GB:AE000932; GB:AE000666; NID
A;Experimental source: strain Delta H
C;Genetics:
A;Gene: MTH1770
C;Complex: dodecamer; double hexamer [validated, MUID:20144074]
C;Function: <HEL>
A;Description: EC 3.6.1.-; DNA helicase; ATP-dependent DNA unwinding activity [validated
C;Function: <ATP>
A;Description: EC 3.6.1.3 [validated, MUID:20144074]; adenosinetriphosphatase, DNA-depen
C;Function: <DNA>
A;Description: ATP-independent single-stranded DNA-binding; ATP-dependent double-strande
C;Superfamily: Archaeoglobus replication licensing factor MCM-type homolog; MCM homology
C;Keywords: cell cycle control; DNA replication initiation; hydrolase
F;98-570/Domain: MCM homology <MCM>

  Query Match             22.1%;  Score 1020.5;  DB 1;  Length 666;
  Best Local Similarity   34.2%;  Pred. No. 1.2e-49;
  Matches  243;  Conservative  138;  Mismatches  244;  Indels   85;  Gaps   16;

Qy        187 DVLRAIRRF-----LHNYRSSAHDLNSKYIQIIEETVEREEDTLNIDMSDIYDHDPDLYA 241
              | : : :|      | :|:    :    ||   :            :: :|  |:   ||||
Db          6 DKSKTLTKFEEFFSLQDYKDRVFEAIEKYPNV---------RSIEVDYLDLEMFDPDLAD 56

Qy        242 KIVRYPLDIIPLLDTECQEVATSLLPTFEKHIEARPFNLKAS-----VHMRELNPSDIDK 296
               ::  | |:|           :|   :      |:::     |:: |       : :|||     ||
Db         57 LLIEKPDDVI-----RAAQQAIRNIDRLRKNVD---LNIRFSGISNVIPLRELRSKFIGK 108

Qy        297 LVSVKGMVIRCSSIIPEIKGAFFKCLVC--GHSPPLVTVVKGRVEEPTRCEKPECAARNA 354
              |:| |:|  :      | |  |  | |:|  |     |:    ||       : ||: |     ||  | :
Db        109 FVAVDGIVRKTDEIRPRIVKAVFECRGCMRHHA---VTQSTNMITEPSLCS--ECGGR-S 162

Qy        355 MSLIHNRCTFANKQIVRLQETPDAIPEGETPHTVSMCLYNTMVDAVKPGDRIEVTGTVFKA 414
              |: :      |  :  | ::|||   :  :   ||   :::  |  : :||  :  |||  :  |||  :
Db        163 FRLLQDESEFLDTQTLKLQEPLENLSGGEQPRQITVVLEDDLVDTLTPGDIVRVTGTLR- 221

Qy        415 MAVRVGPNQRTLRALYKTYIDCVHVKKSDRGRLQTEDPMEMDKENDMYAGYHESDTSEAA 474
               ||      :|      :|   | : :               |   : ::: :|::
Db        222 -TVRDERTKRFKNFIYGNYTEFLE---------QEFEELQISEEDE------------- 257

Qy        475 NEAKIQKLKELSKLPGIYDRLSRSLAPSIWELEDIKKGLLCQLFGGKAKKIPSGASFRGD 534
                   :|:|||:  |  ||::: ||  ||||      ::|: :   |||||   |::          |||
Db        258 -----EKIKELAGDPNIYEKIIRSTAPSIHGYREVKEAIALQLFGGTGKELDDKTRLRGD 312

Qy        535 INVLLVGDPGTSKSQLLQYVHKIAPRGIYTSGRGSSAVGLTAYVTKDPETRETVLESGAL 594
              |::|:|||||  |||:|:|| |:||||||||:|:| |||||     :| |        ||:|||
Db        313 IHILIVGDPGIGKSQMLKYVSKLAPRGIYTSGKGTSGVGLTAAAVRD-EFGGWSLEAGAL 371

Qy        595 VLSDRGICCIDEFDKMSDNARSMLHEVMEQQTVSVAKGGIIASLNARTSVLACANPSGSR 654
              || |:  |:|| |||  |   ||  :|  :|||||:|  |:|| |:|:|:|  ||||  |||       |
Db        372 VLGDKGNVCVDELDKMREEDRSAIHEALEQQTISIAKAGIMATLNSRCSVLAAANPKFGR 431

Qy        655 YNARLSVIDNIQLPPTLLSRFDLIYLMLDKPDEQNDRRLARHLVALHYENYEVSKQDALD 714
              :::  |: :  | ||  |:||||||||:::  |||||: || ||||:    |::          :|
Db        432 FDSYKSIAEQIDLPSTILSRFDLIFVVEDKPDEEKDRELARHILKTHKEDHMPFE---ID 488

Qy        715 LQTLTAYITYARQHVHPTLSDEAAEDLINGYVEMRQKGNFPGSSKKVITATPRQLESMIR 774
               :  |  || |||::|  | |:|||  :  |  :  ||  ||         :   |  ||||:::|
Db        489 PELLRKYIAYARKNVRPVLTDEAMQVLEDFYVSMRASA---ADEDSPVPITARQLEALVR 545

Qy        775 ISEALARMRFSEVVEKVDAAEAVRLLDVALQQSATDHATGTIDMDLITTGVSASERIRRA 834
              :|||  |:::  |  ||   ||  :|::|     |:|     |   || ||:|  :       |||  :
Db        546 LSEASAKIKLKEHVEAEDARKAIKLSQACLKQVGYDPPETGKIDIDKVEGRTPKSERDKFR 605

Qy        835 NLLAALRE-------------LIADKISPGSSSGLKTSQLLEDIRSQSSV 871
                 ||   ::|              || :  :     :  |  |  :|:  :: : ::
Db        606 LLLELIKEYEDDYGGRAPTNILITEMMDRYNVSEEKVEELIRILKDKGAI 655

RESULT 12
T39991
minichromosome maintenance protein mcm7p [imported] - fission yeast (Schizosaccharomyces
N;Alternate names: cell division control protein 47 homolog
C;Species: Schizosaccharomyces pombe
C;Date: 03-Dec-1999 #sequence_revision 03-Dec-1999 #text_change 09-Jul-2004
C;Accession: T39991; T45282
R;Lyne, M.; Rajandream, M.A.; Barrell, B.G.; Whithead, S.; Chillingworth, T.; Churcher,
submitted to the EMBL Data Library, July 1998
A;Reference number: Z21897
A;Accession: T39991
A;Status: preliminary; translated from GB/EMBL/DDBJ
A;Molecule type: DNA
A;Residues: 1-760 <LYN>
A;Cross-references: UNIPROT:O75001; UNIPARC:UPI000012ED9F; EMBL:AL031158; PIDN:CAA20099.
A;Experimental source: strain 972h-; cosmid c25D12
R;Liang, D.T.; Forsburg, S.L.
submitted to the EMBL Data Library, June 1998
A;Description: Fission yeast mcm7+ is an essential gene required for normal DNA replicat
A;Reference number: Z22955
A;Accession: T45282
A;Status: preliminary; translated from GB/EMBL/DDBJ
A;Molecule type: DNA
A;Residues: 1-760 <LIA>
A;Cross-references: UNIPARC:UPI000012ED9F; EMBL:AF070481; PIDN:AAC23693.1
A;Experimental source: strain Sp.011
C;Genetics:
A;Gene: SPBC25D12.03c; mcm7
A;Map position: 2
C;Function:
A;Description: essential replication factor
C;Superfamily: replication licensing factor MCM7; MCM homology
F;170-662/Domain: MCM homology <MCM>

  Query Match             22.1%;  Score 1016.5;  DB 2;  Length 760;
  Best Local Similarity   33.2%;  Pred. No. 2.5e-49;
  Matches  248;  Conservative  147;  Mismatches  231;  Indels  121;  Gaps   22;

Qy        201 SSAHD---LNSKYIQIIEETVEREEDTLNIDMSDIYDHDP-----------------DLY 240
              | | |    | |||: |:::    || : :|:|::|:|: ||                    :|:
Db         44 SDAGDEPFLKSKYMDILQKISNRESNVINVDLNDLYEFDPSDTQLLHNIESNAKRFVELF 103

Qy        241 AKI-----------VRYPLDIIPL-----------LDTECQEVATSLLPTFEKHI----- 273
              ::                :  |  ::: :               :| |   :     |    :: :
Db        104 SQCADALMPPPTVEINYRNEVLDVIMQQRVQRNENIDPEHKGFPPELTRGYDLYFRPVTR 163

Qy        274 EARPFNLKASVHMRELNPSDIDKLVSVKGMVIRCSSIIPEIKGAFFKCLVCGHSPPLVTV 333
                :||:         :|:|    :: |::|:|:| |  | |  :        :  |   ||:     :
Db        164 NKKPFS------VRDLRGENLGSLLTVRGIVTRTSDVKPSLTVNAYTCDRCGYE--VFQE 215

Qy        334 VKGRVEEP-TRCEKPECAARNAMSLIHNRCTFANK----QIVRLQETPDAIPEGETPHTV 388
              :: :     | :  |   ||    :|    :       | |:|      | |::||   : :|  |   | ::
Db        216 IRQKTFLPMSECPSDECKKNDAKGQLF-MSTRASKFLPFQEVKIQELTNQVPIGHIPRSL 274

Qy        389 SMCLYNTMVDAVKPGDRIEVTGTVFKA------MAVRVGPNQRTLRALYKTYIDCVHVKKS 442
              :: ||   :   :| ||| ::::|:|               |:| |               |  ||::| :| :
Db        275 TVHLYGAITRSVNPGDIVDISGIFLPTPYTGFRAMRAG-------LLTDTYLECHYVSQI 327

Qy        443 DRGRLQTEDPMEMDKENDMYAGYHESDTSEAANEAKIQKLKELSKLPGIYDRLSRSLAPS 502
               :                       |     ::   ||||          : ||::    :|::|::|:||
Db        328 IKN---------------YTNIEKTPQSEAA-------IAELNQGGNVYEKLAKSIAPE 364

Qy        503 IWELEDIKKGLLCQLFGGKAKKIPSGASFRGDINVLLVGDPGTSKSQLLQYVHKIAPRGI 562
              |:   ||:|| ||   |   ||   |::   |        |||||:  | ||||  :|||||:|:  |:||||:
Db        365 IYGHEDVKKALLLLLVGGVTKELGDGMRIRGDINICLTGDPGVAKSQLLKYISKVAPRGV 424

Qy        563 YTSGRGSSAVGLTAYVTKDPETRETVLESGALVLSDRGICCIDEFDKMSDNARSMLHEVM 622
```

```
              |: :  |:| |||||:||:||| |::| |  || :|  || ||  ||:|||||  :
Db        722 -VGAGRGQISAYPRQLESLIRLSEAHAKVRLSNQVELLDVEEAWRLHREALKQSATDPLS 780

Qy        814 GTIDMDLITTGVSASERIRRANLLAALRELIADKISPGSSSGLKTSQLLEDIRSQSSVDV 873
              | ||: ::|||:| : | :||:|:||::| :  |   |    :   :|  ||:  | : :
Db        781 GKIDVGILTTGLSTAARKKRADLVAAIKENLKKK---GKVLTVPYQKLFSDIKEGSQIMI 837

Qy        874 SLQDIKNALGSLQGEGFLTVHGDIVKRV 901
              : :  ::||  :| || : | |    |:
Db        838 TREQFEDALKEVQDEGAIVVMGKNTIRI 865


RESULT 9
T50408
cdc21 protein [imported] - fission yeast (Schizosaccharomyces pombe) (fragment)
C;Species: Schizosaccharomyces pombe
C;Date: 09-Jun-2000 #sequence_revision 09-Jun-2000 #text_change 02-Sep-2000
C;Accession: T50408
R;Seeger, K.; Harris, D.; Wood, V.; Rajandream, M.A.; Barrell, B.G.
submitted to the EMBL Data Library, February 2000
A;Reference number: Z25039
A;Accession: T50408
A;Status: preliminary; translated from GB/EMBL/DDBJ
A;Molecule type: DNA
A;Residues: 1-407 <SEE>
A;Cross-references: UNIPARC:UPI00001690AB; EMBL:AL157991; PIDN:CAB76210.1; GSPDB:GN00068
A;Experimental source: strain 972h(-); cosmid c24B10
C;Genetics:
A;Gene: SPDB:SPCC24B10.01
A;Map position: 3
C;Superfamily: replication licensing factor MCM4; MCM homology
F;1-301/Domain: MCM homology (fragment) <MCM>

  Query Match             23.1%;  Score 1064;  DB 2;  Length 407;
  Best Local Similarity   55.9%;  Pred. No. 2.1e-52;
  Matches  221;  Conservative   66;  Mismatches  100;  Indels    8;  Gaps    4;

Qy        503 IWELEDIKKGLLCQLFGGKAKKIPSGAS--FRGDINVLLVGDPGTSKSQLLQYVHKIAPR 560
              |:|::|:|||| |||||      |||     :|||||:|: ||| |||||:||||||||
Db          1 IYEMDDVKKGLLLQLFGGTNKSFHKGASPRYRGDINILMCGDPSTSKSQILKYVHKIAPR 60

Qy        561 GIYTSGRGSSAVGLTAYVTKDPETRETVLESGALVLSDRGICCIDEFDKMSDNARSMLHE 620
              |:||||:||||||||||||:|:| :|:: |||||||||| |||||||||||||  ||:|||
Db         61 GVYTSGKGSSAVGLTAYITRDQDTKQLVLESGALVLSDGGICCIDEFDKMSDATRSILHE 120

Qy        621 VMEQQTVSVAKGGIIASLNARTSVLACANPSGSGSRYNARLSVIDNIQLPPTLLSRFDLIYL 680
              |||||||:||| |||  :||||||:|| ||| ||:||  |  |  ||||||||||||:||
Db        121 VMEQQTVTVAKAGIITTLNARTSILASANPIGSKYNPDLPVTKNIDLPPTLLSRFDLVYL 180

Qy        681 MLDKPDEQNDRRLARHLVALHYEN--YEVSKQDALDLQTLTAYITYARQHVHPTLSDEAA 738
              :||: ||  ||:|| |:|::: |:    :  :   :: ||:|||||| :::| :|:|||
Db        181 ILDRVDETLDRKLANHIVSMYMEDTPEHATDMEVFSVEFLTSYITYARNNINPVISEEAA 240

Qy        739 EDLINGYVEMRQKGNFPGSSKKVITATPRQLESMIRISEALARMRFSEVVEKVDAAEAVR 798
              ::|:| || ||: |   :|:| |||| |||||||||:||| |:|  |||  |  ||
Db        241 KELVNAYVGMRKLGEDVRASEKRITATTRQLESMIRLSEAHAKMHLRNVVEVGDVLEAAR 300

Qy        799 LLDVALQQSATDHATGTIDMDLITTGVSASERIRRANLLAALRELIADKISPGSSSGLKT 858
              |:  |::  ||| |||  | :|||  |:   |:    :::   | ||::      |  :
Db        301 LIKTAIKDYATDPATGKISLDLIY--VNERETLVPEDMVKELANLISNLTVGGKT--MLV 356

Qy        859 SQLLEDIRSQSSVDVSLQDIKNALGSLQGEGFLTV 893
              ||||    | |||  :    | :  ||:|:  | : |
Db        357 SQLLTRFREQSSTRLDASDFEACLGALERRGRIKV 391


RESULT 10
H90226
minichromosome maintenance protein MCM [imported] - Sulfolobus solfataricus
C;Species: Sulfolobus solfataricus
C;Date: 24-May-2001 #sequence_revision 24-May-2001 #text_change 09-Jul-2004
C;Accession: H90226
R;She, Q.; Singh, R.K.; Confalonieri, F.; Zivanovic, Y.; Allard, G.; Awayez, M.J.; Chan-
Jong, I.; Jeffries, A.C.; Kozera, C.J.; Medina, N.; Peng, X.; Thi-Ngoc, H.P.; Redder, 
arrett, R.A.; Ragan, M.A.; Sensen, C.W.; Van der Oost, J.
submitted to GenBank, April 2001
A;Description: Sulfolobus solfataricus complete genome.
A;Reference number: A99139
A;Accession: H90226
A;Status: preliminary
A;Molecule type: DNA
A;Residues: 1-686 <KUR>
A;Cross-references: UNIPROT:Q9UXG1; UNIPARC:UPI0000066A8C; GB:AE006641; NID:g13813947; 
C;Genetics:
A;Gene: SSO0774

  Query Match             22.5%;  Score 1039;  DB 2;  Length 686;
  Best Local Similarity   34.4%;  Pred. No. 1.2e-50;
  Matches  244;  Conservative  150;  Mismatches  252;  Indels   64;  Gaps   15;

Qy        195 FLHNYRSSAHDLNSKYIQIIEETVEREEDTLNIDMSDIYDHDPDLYAKIVRYPLDIIPLL 254
              ||  :: : :    :|||: | |   : :| |: ||:   : :|  :|:     |:|:|
Db         17 FLTTFKGNNN--QNKYIERINELVAYRKKSLIIEFSDVLSFNENLAYEIINNTKIILPIL 74

Qy        255 DTECQEVATSLLPTFEKHIE---ARPFNLKASVHMRELNPSDIDKLVSVKGMVIRCSSII 311
               : :   :   ||::: ||      |      :  :  :|| ||::: |:::: : :
Db         75 EGALYDHILQLDPTYQRDIEKVHVRIVGIPRVIELRKIRSTDIGKLITIDGILVKVTPVK 134

Qy        312 PEIKGAFFK-----CLVCGHSP-----PLVTVVKGRVEEPTRCEKPECAARNAMSLIHNR 361
               |  | :|     |:       |   | |      :| || |  |:|       ||  :
Db        135 ERIYKATYKHIHPDCMQEFEWPEDEEMPEV------LEMPTIC--PKCGKPGQFRLIPEK 186

Qy        362 CTFANKQIVRLQETPDAIPEGETPHTVSMCLYNTMVDAVKPGDRIEVTGVFKAMAVRVGP 421
                 : |  :|| |: :| | |: |  : : | : :||: :||||::|||:      :  |
Db        187 TKLIDWQKAVIQERPEEVPSGQLPRQLEIILEDDLVDSARPGDRVKVTGILDIK--QDSP 244

Qy        422 NQRTLRALYKTYIDCVHVKKSDRGRLQTEDPMEMDKENDMYAGYHESDTSEAANEAKIQK 481
               :|  ||::  |:      :: | :        | : : :|::                :|
Db        245 VKRGSRAVFDIYMKVSSIEVSQK----VLDEVIISEEDE------------------KK 281

Qy        482 LKELSKLPGIYDRLSRSLAPSIWELEDIKKGLLCQLFGGKAKKIPSGASFRGDINVLLVG 541
              :|:|:|  |  |  ||:  |:||:  ::|:  |  ||||  |:  ||  ||  |:|::|
Db        282 IKDLAKDPWIRDRIISSIAPSIYGHWELKEALALALFGG-VPKVLEDTRIRGDIHILIIG 340

Qy        542 DPGTSKSQLLQYVHKIAPRGIYTSGRGSSAVGLTAYVTKDPETRETVLESGALVLSDRGI 601
              ||||:|||:||:: ::|||  :||:|:||:| ||||  | ::   |  |  ||:|||||:| ||
Db        341 DPGTAKSQMLQFISRVAPRAVYTTGKGSTAAGLTAAVVREKGTGEYYLEAGALVLADGGI 400

Qy        602 CCIDEFDKMSDNARSMLHEVMEQQTVSVAKGGIIASLNARTSVLACANPSGSGSRYNARLSV 661
                ||| |||  |   :|| ||||||:|| ||:| ||||  :|:| ||   ||  :  |
Db        401 AVIDEIDKMRDEDRVAIHEAMEQQTVSIAKAGIVAKLNARAAVIAAGNPKFGRYISERPV 460

Qy        662 IDNIQLPPTLLSRFDLIYLMLDKPDEQNDRRLARHLVALHYENYEVSKQDALDLQTLTAY 721
               ||| ||||:||||||||::: |:| || || || ::: :|        | :: :|: || |
Db        461 SDNINLPPTILSRFDLIFILKDQPGEQ-DRELANYILDVHSGK---STKNIIDIDTLRKY 516

Qy        722 ITYARQHVHPTLSDEAAEDLINGYVEMRQKGNFPGSSKKVITATPRQLESMIRISEALAR 781
               | |||::| |  :: |   : : :||||:|   :     |      ||||||::||||| |: 
Db        517 IAYARKYVTPKITSEAKNLITDFFVEMRKKSSETPDSP--ILITPRQLEALIRISEAYAK 574

Qy        782 MRFSEVVEKVDAAEAVRLLDVALQQSATDHATGTIDMDLITTGVSASERIRRANLLAALR 841
                    |  : ||  |: : :  :  |:    | : :|||:| | ||    | |  |  :   ::  :
Db        575 MALKAEVTREDAERAINIMRLFLESVGVDMESGKIDIDTIMTGKPKSAREKMMKIIEIID 634

Qy        842 ELIADKISPGSSSGLKTSQLLEDIRSQSSVDVSLQDIKNALGSLQGEGFL 891
               |       ||  |     :|:: :    | :    :|:  |  ::  | :
Db        635 SLAV------SSECAKVKDILKEAQ---QVGIEKSNIEKLLTDMRKSGII 675


RESULT 11
D69103
DNA helicase (EC 3.6.1.-) MTH1770 [validated] - Methanobacterium thermoautotrophicum (s
N;Alternate names: cell division control protein CDC21/CDC54 homolog; replication licen
N;Contains: adenosinetriphosphatase (EC 3.6.1.3)
C;Species: Methanobacterium thermoautotrophicum
C;Date: 13-Aug-1999 #sequence_revision 13-Aug-1999 #text_change 09-Jul-2004
```

```
  Query Match             38.7%;  Score 1784.5;  DB 2;  Length 858;
  Best Local Similarity   46.1%;  Pred. No. 2.3e-92;
  Matches  387;  Conservative  142;  Mismatches  238;  Indels   73;  Gaps   17;

Qy         78 NSVYSSPYDAGTPGTPGTPVATPVYATPVGTPMGTP-SFHRGTPQYKQRSELGSQGK--- 133
Db         72 HSAHQSELDLSSPLT---------YGTPSSRVEGTPRSGIRGTPA-RQRADLGSARKVKQ 121

Qy        134 -PLHRRRRSQSREPGHRSPSREPSADGRPSESAEPDDTLGGEYAYVWGTNVNIPDVLRAI 192
Db        122 VDLH---------------SDQPAAE----ELVTSEQSL-GQKLVIWGTDVNVAICKEKF 161

Qy        193 RRFLHNY--------RSSAHDLNSK-YIQIIEETVEREEDTLNIDMSDIYDHDPDLYAKI 243
Db        162 QRFVQRFIDPLAKEEENVGLDLNEPIYMQRLEEINVVGEPFLNIDCDHLRSFDQDLYRQL 221

Qy        244 VRYPLDIIPLLDTECQEVATSLLP--TFEKHIEARPFNLKASVHMRELNPSDIDKLVSVK 301
Db        222 VCYPQEVIPTFDMAANEIFFERYPDSILEHQIQVRPYNALKTRNMRSLNPEDIDQLITIS 281

Qy        302 GMVIRCSSIIPEIKGAFFKCLVCGHSPPLVTVVKGRVEEPTRCEKPECAARNAMSLIHNR 361
Db        282 GMVIRTSQIIPEMQESFFKCQVCAFT-TRVEIDRGRIAEPSVCK--HCNTTHSMALIHNR 338

Qy        362 CTFANKQIVRLQETPDAIPEGETPHTVSMCLYNTMVDAVKPGDRIEVTGVFKAMAVRVGP 421
Db        339 SMFSDKQMIKLQESPGDMPAGQTPHTTILYAHNDLVDKVQPGDRVNVTGIYRAVPIRVNP 398

Qy        422 NQRTLRALYKTYIDCVHVKKSDRGRLQTEDPMEMDKENDMYAGYHESDTSEAANEAKIQK 481
Db        399 RVRNVKSVYKTHIDVIHYRKTDSKRLHGIDE---DTEQKMF------------TEERVAV 443

Qy        482 LKELSKLPGIYDRLSRSLAPSIWELEDIKKGLLCQLFGGKAKKI--PSGASFRGDINVLL 539
Db        444 LKELAAKPDIYERLAAALAPSIYEHEDIKKGILLQLFGGTRKDFSHTGRGKFRAEVNILL 503

Qy        540 VGDPGTSKSQLLQYVHKIAPRGIYTSGRGSSAVGLTAYVTKDPETRETVLESGALVLSDR 599
Db        504 CGDPGTSKSQLLQYVYNLVPRGQYTSGKGSSAVGLTAYVMKDPETRQLVLQTGALVLSDN 563

Qy        600 GICCIDEFDKMSDNARSMLHEVMEQQTVSVAKGGIIASLNARTSVLACANPSGSRYNARL 659
Db        564 GICCIDEFDKMNESTRSVLHEVMEQQTLSIAKAGIICQLNARTSVLAAANPVESQWNPKK 623

Qy        660 SVIDNIQLPPTLLSRFDLIYLMLDKPDEQNDRRLARHLVALHYENYEVSKQDALDLQTLT 719
Db        624 TTIENIQLPHTLLSRFDLIFLMLDPQDETYDRRLAHHLVVLYYQSEEQLKEEHLDMAVLK 683

Qy        720 AYITYARQHVHPTLSDEAAEDLINGYVEMRQKGNFPGSSKKVITATPRQLESMIRISEAL 779
Db        684 DYIAYARTYVNPRLGEEASQALIEAYVDMRK----IGSGRGMVSAYPRQLESLIRLSEAH 739

Qy        780 ARMRFSEVVEKVDAAEAVRLLDVALQQSATDHATGTIDMDLITTGVSASERIRRANLLAA 839
Db        740 AKVRFSSKVETIDVEEAKRLHREALKQSATDPRTGIVDISILTTGMSATARKRKEELAQV 799

Qy        840 LRELIADKISPGSSSGLKTSQLLEDIRSQSSVDVSLQDIKNALGSLQGEGFLTVHGDIVK 899
Db        800 LKKLIQSK---GKTPAFKYQQLFEDLRGQSDAAITKDMFDEALHALADEDYLTVTGKTVR 856
```

RESULT 8
S59872
replication licensing factor MCM4 - fruit fly (Drosophila melanogaster)
N;Alternate names: cell division cycle control protein CDC21/CDC54; gene disc proliferat
C;Species: Drosophila melanogaster
C;Date: 13-Aug-1999 #sequence_revision 13-Aug-1999 #text_change 09-Jul-2004
C;Accession: S59872
R;Feger, G.; Vaessin, H.; Su, T.T.; Wolff, E.; Jan, L.Y.; Jan, Y.N.
EMBO J. 14, 5387-5398, 1995
A;Title: dpa, a member of the MCM family, is required for mitotic DNA replication but nc
A;Reference number: S59872; MUID:96080174; PMID:7489728
A;Accession: S59872
A;Molecule type: mRNA
A;Residues: 1-866 <FEG>
A;Cross-references: UNIPROT:Q26454; UNIPARC:UPI000016C029; GB:S80255; NID:g1245869; PIDN
C;Comment: The complex of six MCM proteins is one of several proteins that must be bound
phosphorylated and dissociate from the chromatin.
C;Genetics:
A;Gene: dpa; disc proliferation abnormal
A;Cross-references: FlyBase:FBgn0015929
C;Complex: The predominant form is a heterohexamer of MCM2, MCM3, MCM4, MCM5, MCM6, and
C;Function:
A;Description: part of the replication licensing system that permits DNA replication to
C;Superfamily: replication licensing factor MCM4; MCM homology
C;Keywords: cell cycle control; DNA replication initiation; heterohexamer; nucleus; phos
F;274-766/Domain: MCM homology <MCM>

```
  Query Match             36.7%;  Score 1689.5;  DB 1;  Length 866;
  Best Local Similarity   41.5%;  Pred. No. 5.1e-87;
  Matches  385;  Conservative  161;  Mismatches  281;  Indels  101;  Gaps   19;

Qy         12 VSSPYPSQSEGVSTPLPQVTSPSFDNAASPVAGRRAVRQTPTSAVRRRGRET-------D 64
Db          1 MSSPARSPSVGGATPKQGARTPTRGIASQDV--ETPMRMGPGRAVRPSDNISLPPTSPGN 58

Qy         65 SARRRRSRSRSLGNSV----YSSPYDAGTPGTPGTPVATPVYATPVGTPMGTPSFHRGTP 120
Db         59 ISLPATSPARGLGANMSEIDLSSPLNYGTPSSMGS-IRTP------------------- 97

Qy        121 QYKQRSELGSQGKPLHRRRRSQSREPGHRSPSR------------EPSADGRPSESAEP- 167
Db         98 ----RS--GIRGTPLRAR-------PDIRTDKRIRQVAIGGGSGLEPIPE-KGSETTDPV 143

Qy        168 DDTLGGEYAYVWGTNVNIPDVLRAIRRFLHNY---------RSSAHDLNSK-YIQIIEET 217
Db        144 SESSQAPQLVVWGTNVVVSQCKSKFKSFIMRFIDPSAEQDEISENIDVNQPLYLQKLEEI 203

Qy        218 VEREEDTLNIDMSDIYDHDPDLYAKIVRYPLDIIPLLDTECQEVATSLLPT--FEKHIEA 275
Db        204 HTLEEPYLNLNCAHLKTFDQDLYRQLICYPQEVIPGFDMAINEMFFERYPAALLEHQIQV 263

Qy        276 RPFNLKASVHMRELNPSDIDKLVSVKGMVIRCSSIIPEIKGAFFKCLVCGHSPPLVTVVK 335
Db        264 RPFNADKTRNMRSLNPEDMDQLISISGMVIRSSNVIPEMREAFFSCNICSFS-TTVEVDR 322

Qy        336 GRVEEPTRCEKPECAARNAMSLIHNRCTFANKQIVRLQETPDAIPEGETPHTVSMCLYNT 395
Db        323 GRINQPTLC--TNCNTNHCFRLIHNRSEFTDKQLVKLQESPDDMAAGQTPHNVLLYAHND 380

Qy        396 MVDAVKPGDRIEVTGVFKAMAVRVGPNQRTLRALYKTYIDCVHVKKSDRGRLQTEDPMEM 455
Db        381 LVDKVQPGDRVTVTGIYRATPLKTGGLSSSVKSVYKTHVDVVHFRKVDNKRLYEE----- 435

Qy        456 DKENDMYAGYHESDTSEAANEAKIQKLKELSKLPGIYDRLSRSLAPSIWELEDIKKGLLC 515
Db        436 -----------EEGKDHIFPPERVELLQLLAKKPDIYDRLARAIAPSIYENDDIKKGILL 484

Qy        516 QLFGGKAKKIPS--GASFRGDINVLLVGDPGTSKSQLLQYVHKIAPRGIYTSGRGSSAVG 573
Db        485 QLFGGTKKKHATLGRQNFRSEIHLLLCGDPGTSKSQMLQYVFNLVPRSQYTSGRGSSAVG 544

Qy        574 LTAYVTKDPETRETVLESGALVLSDRGICCIDEFDKMSDNARSMLHEVMEQQTVSVAKGG 633
Db        545 LTAYVTKDPETRQLVLQTGALVLADNGVCCIDEFDKMNDSTRSVLHEVMEQQTLSIAKAG 604

Qy        634 IIASLNARTSVLACANPSGSRYNARLSVIDNIQLPPTLLSRFDLIYLMLDKPDEQNDRRL 693
Db        605 IICQLNARTSILAAANPAESQWNKRKNIIDNVQLPHTLLSRFDLIFLVLDPQDEIFDKRL 664

Qy        694 ARHLVALHYENYEVSKQDALDLQTLTAYITYARQHVHPTLSDEAAEDLINGYVEMRQKGN 753
Db        665 ASHLVSLYYVTRHEEDTMFDMSVLRDYIAYAREHLSPTLSDEAQQRLIQAYVDMRK--- 721

Qy        754 FPGSSKKVITATPRQLESMIRISEALARMRFSEVVEKVDAAEAVRLLDVALQQSATDHAT 813
```

```
Db        622 SQWNPKKTTIENIQLPHTLLSRFDLIFLMLDPQDEAYDRRLAHHLVALYYQSEEQAEEEL 681

Qy        713 LDLQTLTAYITYARQHVHPTLSDEAAEDLINGYVEMRQKGNFPGSSKKVITATPRQLESM 772
              ||:  |  || ||   : | ||:||:: ||  ||:||:       |||: :::| ||||||:
Db        682 LDMAVLKDYIAYAHSTIMPRLSEEASQALIEAYVDMRK----IGSSRGMVSAYPRQLESL 737

Qy        773 IRISEALARMRFSEVVEKVDAAEAVRLLDVALQQSATDHATGTIDMDLITTGVSASERIR 832
              ||::|| |::| |  || :|  || ||    ||:||||| || :|: ::|||:||: | |
Db        738 IRLAEAHAKVRLSNKVEAIDVEEAKRLHREALKQSATDPRTGIVDISILTTGMSATSRKR 797

Qy        833 RANLLAALRELIADKISPGSSSGLKTSQLLEDIRSQSSVDVSLQDIKNALGSLQGEGFLT 892
               :  |  ||::||    :| | :   ||  || |||| || : ::    : || :|  : |||
Db        798 KEELAEALKKLI---LSKGKTPALKYQQLFEDIRGQSDIAITKDMFEEALRALADDDFLT 854

Qy        893 VHGDIVK 899
              | |  |:
Db        855 VTGKTVR 861
```

RESULT 6
S64720
replication licensing factor MCM4 - African clawed frog
N;Alternate names: cell division cycle control protein CDC21/CDC54
C;Species: Xenopus laevis (African clawed frog)
C;Date: 13-Aug-1999 #sequence_revision 13-Aug-1999 #text_change 09-Jul-2004
C;Accession: S64720; S26643; S25529
R;Coue, M.; Kearsey, S.E.; Mechali, M.
EMBO J. 15, 1085-1097, 1996
A;Title: Chromatin binding, nuclear localization and phosphorylation of Xenopus cdc21 ar
A;Reference number: S64720; MUID:96183193; PMID:8605878
A;Accession: S64720
A;Status: nucleic acid sequence not shown; translation not shown
A;Molecule type: mRNA
A;Residues: 1-863 <COU>
A;Cross-references: UNIPROT:P30664; UNIPARC:UPI000012ED9C; EMBL:U29178; NID:g1002597; PI
A;Note: the nucleotide sequence was submitted to the EMBL Data Library, June 1995
R;Coxon, A.; Maundrell, K.; Kearsey, S.E.
Nucleic Acids Res. 20, 5571-5577, 1992
A;Title: Fission yeast cdc21(+) belongs to a family of proteins involved in an early ste
A;Reference number: S26640; MUID:93087163; PMID:1454522
A;Accession: S26643
A;Molecule type: DNA
A;Residues: 513-523,'Y',525-538,'G',540-553,'G',555-588 <COX>
A;Cross-references: UNIPARC:UPI000017151B; EMBL:Z15033; NID:g64612; PIDN:CAA78751.1; PID
C;Comment: The complex of six MCM proteins is one of several proteins that must be bound
 phosphorylated and dissociate from the chromatin.
C;Genetics:
A;Gene: cdc21
C;Complex: The predominant form is a heterohexamer of MCM2, MCM3, MCM4, MCM5, MCM6, and
C;Function:
A;Description: part of the replication licensing system that permits DNA replication to
C;Superfamily: replication licensing factor MCM4; MCM homology
C;Keywords: cell cycle control; DNA replication initiation; heterohexamer; nucleus; phos
F;271-764/Domain: MCM homology <MCM>

```
  Query Match             39.1%;  Score 1801.5;  DB 1;  Length 863;
  Best Local Similarity   43.7%;  Pred. No. 2.6e-93;
  Matches  398;  Conservative  147;  Mismatches  249;  Indels  116;  Gaps   18;

Qy         51 TPTSAVRRRGRETD-------------SARRRRSRSRSLG-------------------- 77
              ||:    :||| ::             | |||    | |:|
Db          7 TPSRRRNKRGRGSNPPTPHGEEVQSPPSQRRRTEDSTSIGELLPMPTSPSGDVQSPSGQE 66

Qy         78 ----------NSVYSSPYDAGTPGTPGTPVATPVYATPVGTPMGTP-SFHRGTPQYKQRS 126
                        :| :  |  :| |             | ||         ||| |  |||| :||
Db         67 LLFSSPVPSRHSAHQSELDLSSPLT---------YGTPSSRVEGTPRSGIRGTPA-RQRP 116

Qy        127 ELGSQGK----PLHRRRRSQSREPGHRSPSREPSADGRPSESAEPDDTLGGEYAYVWGTN 182
              :|||    |        ||                | :|:|:    |      : :| |:   :|||:
Db        117 DLGSARKVKQVDLH---------------SDQPAAE----ELVTSEQSL-GQKLVIWGTD 156

Qy        183 VNIPDVLRAIRRFLHNYRSSAH---------DLNSK-YIQIIEETVEREEDTLNIDMSDIY 233
              ||:        :||:  :   :              |||    |:| :||        :  ||||    :

Db        157 VNVATCKEKFQRFVQRFIDPSAKEEDNVGLDLNEPIYMQRLEEINVVGDPFLNIDCDHLR 216

Qy        234 DHDPDLYAKIVRYPLDIIPLLDTECQEVATSLLP--TFEKHIEARPFNLKASVHMRELNP 291
              : | ||| ::| || ::|| |    |:     |    | |: ||:|    : :|| |||
Db        217 NFDQDLYRQLVCYPQEVIPTFDMAANEIFFERYPDSILEHQIQVRPYNALKTRNMRSLNP 276

Qy        292 SDIDKLVSVKGMVIRCSSIIPEIKGAFFKCLVCGHSPPLVTVVKGRVEEPTRCEKPECAA 351
              |||:|::: ||||| | ||||:: ||||| ||   :    | : :||: ||: |:    |
Db        277 EDIDQLITISGMVIRTSQIIPEMQEAFFKCQVCAFT-TRVEIDRGRIAEPSVCK--HCNT 333

Qy        352 RNAMSLIHNRCTFANKQIVRLQETPDAIPEGETPHTVSMCLYNTMVDAVKPGDRIEVTGV 411
              ::|:||||| |::||:::|||:|: :| |:||||  :   :| :|| |:||||: |||:
Db        334 THSMALIHNRSMFSDKQMIKLQESPEDMPAGQTPHTTILYGHNDLVDKVQPGDRVNVTGI 393

Qy        412 FKAMAVRVGPNQRTLRALYKTYIDCVHVKKSDRGRLQTEDPMEMDKENDMYAGYHESDTS 471
              ::|: :|| |  | ::::|||:|| :| :|:| ||                       | |
Db        394 YRAVPIRVNPRVRNVKSVYKTHIDVIHYRKTDSKRLH---------------GIDEDTEQ 438

Qy        472 EAANEAKIQKLKELSKLPGIYDRLSRSLAPSIWELEDIKKGLLCQLFGGKAKKI--PSGA 529
              :    | ::  ||||:  | ||:||: :|||||:| ||||||:| |||||   |
Db        439 KLFTEERVAMLKELAAKPDIYERLAAALAPSIYEHEDIKKGILLQLFGGTRKDFSHTGRG 498

Qy        530 SFRGDINVLLVGDPGTSKSQLLQYVHKIAPRGIYTSGRGSSAVGLTAYVTKDPETRETVL 589
              || ::|:|| ||||||||||||||   : |||  ||||:||||||||||| ||||||: ||
Db        499 KFRAEVNILLCGDPGTSKSQLLQYVFNLVPRGQYTSGKGSSAVGLTAYVMKDPETRQLVL 558

Qy        590 ESGALVLSDRGICCIDEFDKMSDNARSMLHEVMEQQTVSVAKGGIIASLNARTSVLACAN 649
              ::||||||| ||||||||||||:::: ||:||||||||||:|:|| |||  ||||||||| ||
Db        559 QTGALVLSDNGICCIDEFDKMNESTRSVLHEVMEQQTLSIAKAGIICQLNARTSVLAAAN 618

Qy        650 PSGSRYNARLSVIDNIQLPPTLLSRFDLIYLMLDKPDEQNDRRLARHLVALHYENYEVSK 709
              |  |::| : : |:||||| |||||||||:|||| ||  ||||| ||||:|:: | |
Db        619 PVESQWNPKKTTIENIQLPHTLLSRFDLIFLMLDPQDEAYDRRLAHHLVALYYQSEEQMK 678

Qy        710 QDALDLQTLTAYITYARQHVHPTLSDEAAEDLINGYVEMRQKGNFPGSSKKVITATPRQL 769
              :: ||:   | || ||| :|:| ||:||:: ||  || ||:       || : :::| ||||
Db        679 EEHLDMAVLKDYIAYARTYVNPRLSEEASQALIEAYVSMRK----IGSGRGMVSAYPRQL 734

Qy        770 ESMIRISEALARMRFSEVVEKVDAAEAVRLLDVALQQSATDHATGTIDMDLITTGVSASE 829
              ||:|| :|| |::|||  || :|  || ||    ||:||||| || :|: ::|||:||:
Db        735 ESLIRRAEAHAKVRFSNKVETIDVEEAKRLHREALKQSATDPRTGIVDISILTTGMSATA 794

Qy        830 RIRRANLLAALRELIADKISPGSSSGLKTSQLLEDIRSQSSVDVSLQDIKNALGSLQGEG 889
              | |:  |   |::||  |    | :  ||  || ||:| ||   ::         || :|  :
Db        795 RKRKEELAQVLKKLIQSK---GKTPALKYQQLFEDLRGQSDAAITKDMFDEALHALADDD 851

Qy        890 FLTVHGDIVK 899
              :||| |  |:
Db        852 YLTVTGKTVR 861
```

RESULT 7
T47223
replication licensing factor MCM4 [validated] - African clawed frog
N;Alternate names: cdc21p
C;Species: Xenopus laevis (African clawed frog)
C;Date: 20-Apr-2000 #sequence_revision 20-Apr-2000 #text_change 09-Jul-2004
C;Accession: T47223
R;Kubota, Y.; Mimura, S.; Nishimoto, S.; Masuda, T.; Nojima, H.; Takisawa, H.
EMBO J. 16, 3320-3331, 1997
A;Title: Licensing of DNA replication by a multi-protein complex of MCM/P1 proteins in :
A;Reference number: Z24400; MUID:97357318; PMID:9214647
A;Accession: T47223
A;Status: preliminary; translated from GB/EMBL/DDBJ
A;Molecule type: mRNA
A;Residues: 1-858 <KUB>
A;Cross-references: UNIPROT:O42589; UNIPARC:UPI00000FDD1C; EMBL:U44049; NID:g2231172; P:
C;Complex: the replication licensing complex consists of at least six proteins; MCM2 (P:
alidated, MUID:97357318]
C;Function:
A;Description: MCM4 is a component of the replication licensing factor that permits DNA
C;Superfamily: replication licensing factor MCM4; MCM homology
F;266-759/Domain: MCM homology <MCM>

```
                 :|||:||:          :|||   :             :     |:     |:| : |      |  ||
Db        149 LVIWGTDVNVATCKENFQRFLQCFTDPLAKEEENVGIDITQPLYMQQLGEINITGEPFLN 208

Qy        227 IDMSDIYDHDPDLYAKIVRYPLDIIPLLDTECQEVATSLLP--TFEKHIEARPFNLKASV 284
              ::    |       :|| :::  ||  ::||   |      |:      |      |   |:  ||||     :
Db        209 VNCEHIKSFSKNLYRQLISYPQEVIPTFDMAVNEIFFDRYPDSILEHQIQVRPFNALKTK 268

Qy        285 HMRELNPSDIDKLVSVKGMVIRCSSIIPEIKGAFFKCLVCGHSPPLVTVVKGRVEEPTRC 344
              || ||| |||:|::: ||||| | :|||::  |||:| || |:     | : :||: || |
Db        269 SMRNLNPEDIDQLITISGMVIRTSQLIPEMQEAFFQCQVCAHT-TRVEIDRGRIAEPCSC 327

Qy        345 EKPECAARNAMSLIHNRCTFANKQIVRLQETPDAIPEGETPHTVSMCLYNTMVDAVKPGD 404
                    |    ::|:||||| |::||:::|||:|: :| |:||||: :  :| :|| |:|||
Db        328 --VHCHTTHSMALIHNRSFFSDKQMIKLQESPEDMPAGQTPHTIVLFAHNDLVDKVQPGD 385

Qy        405 RIEVTGVFKAMAVRVGPNQRTLRALYKTYIDCVHVKKSDRGRLQTEDPMEMDKENDMYAG 464
              |: |||:::|: :|| |    ::::|||:|| :| :|:| ||                  |
Db        386 RVNVTGIYRAVPIRVNPRVSNVKSVYKTHIDVIHYRKTDAKRLH---------------G 430

Qy        465 YHESDTSEAANEAKIQKLKELSKLPGIYDRLSRSLAPSIWELEDIKKGLLCQLFGGKAKK 524
                   |    :    :| :::  |||||: | ||:||: :|||||:| ||||||:| ||||| |
Db        431 LDEEAEQKLFSEKRVKLLKELSRKPDIYERLASALAPSIYEHEDIKKGILLQLFGGTRKD 490

Qy        525 I--PSGASFRGDINVLLVGDPGTSKSQLLQYVHKIAPRGIYTSGRGSSAVGLTAYVTKDP 582
                       || :||:||  |||||||||||||||: : |||  ||||:|||||||||||| |||
Db        491 FSHTGRGKFRAEINILLCGDPGTSKSQLLQYVYNLVPRGQYTSGKGSSAVGLTAYVMKDP 550

Qy        583 ETRETVLESGALVLSDRGICCIDEFDKMSDNARSMLHEVMEQQTVSVAKGGIIASLNART 642
              |||: ||:::||||||| |||||||||||:::: ||:||||||||||:|:|| |||  |||||
Db        551 ETRQLVLQTGALVLSDNGICCIDEFDKMNESTRSVLHEVMEQQTLSIAKAGIICQLNART 610

Qy        643 SVLACANPSGSRYNARLSVIDNIQLPPTLLSRFDLIYLMLDKPDEQNDRRLARHLVALHY 702
              |||| |||    |::| : : |:||||| ||||||||||:||||  ||  ||||| |||:|:|
Db        611 SVLAAANPIESQWNPKKTTIENIQLPHTLLSRFDLIFLMLDPQDEAYDRRLAHHLVSLYY 670

Qy        703 ENYEVSKQDALDLQTLTAYITYARQHVHPTLSDEAAEDLINGYVEMRQKGNFPGSSKKVI 762
              :: |  ::: ||:   |  || ||    : | ||:||:: ||  || ||:       |||: ::
Db        671 QSEEQVEEEFLDMAVLKDYIAYAHSTIMPRLSEEASQALIEAYVNMRK----IGSSRGMV 726

Qy        763 TATPRQLESMIRISEALARMRFSEVVEKVDAAEAVRLLDVALQQSATDHATGTIDMDLIT 822
              :| |||||:||::|| ||::|||   || :|  || ||   ||:||||   || :|: ::|
Db        727 SAYPRQLESLIRLAEAHAKVRFSNKVEAIDVEEAKRLHREALKQSATDPRTGIVDISILT 786

Qy        823 TGVSASERIRRANLLAALRELIADKISPGSSSGLKTSQLLEDIRSQSSVDVSLQDIKNAL 882
              ||:||: | |:   ||||:  |    |: :   ||  ||  |||| ||     ::   ||  ||
Db        787 TGMSATSRKRKEELAEALRKLI---LSKGKTPALKYQQLFEDIRGQSDTAITKDMFEEAL 843

Qy        883 GSLQGEGFLTVHGDIVK 899
                :|   : |||| |   |:
Db        844 RALADDDFLTVTGKTVR 860
```

RESULT 5
S65954
replication licensing factor MCM4 - human
N;Alternate names: cell division cycle control protein CDC21/CDC54; S. cerevisiae minich
C;Species: Homo sapiens (man)
C;Date: 13-Aug-1999 #sequence_revision 13-Aug-1999 #text_change 09-Jul-2004
C;Accession: S65954; S43198; S41622
R;Musahl, C.; Schulte, D.; Burkhart, R.; Knippers, R.
Eur. J. Biochem. 230, 1096-1101, 1995
A;Title: A human homologue of the yeast replication protein Cdc21. Interactions with oth
A;Reference number: S65954; MUID:95324568; PMID:7601140
A;Accession: S65954
A;Status: nucleic acid sequence not shown
A;Molecule type: mRNA
A;Residues: 'IRHERVATSSASGRLPRNARRRGRALARSRRERPRCGAGQGSREAGPARACPCRAGTPST',1-863 <MUS>
A;Cross-references: UNIPROT:P33991; UNIPARC:UPI000016AE38; EMBL:X74794; NID:g683749; PID
R;Hu, B.
submitted to the EMBL Data Library, August 1993
A;Reference number: S43198
A;Accession: S43198
A;Molecule type: mRNA
A;Residues: 'ARE',440-702 <HUB1>
A;Cross-references: UNIPARC:UPI000017332D; EMBL:X74794
R;Hu, B.; Burkhart, R.; Schulte, D.; Musahl, C.; Knippers, R.
Nucleic Acids Res. 21, 5289-5293, 1993
A;Title: The P1 family: a new class of nuclear mammalian proteins related to the yeast M
A;Reference number: S41622; MUID:94089373; PMID:8265339
A;Accession: S41622
A;Status: nucleic acid sequence not shown
A;Molecule type: mRNA
A;Residues: 'ARE',440-636 <HUB2>
A;Cross-references: UNIPARC:UPI000017332E; EMBL:X74794
C;Comment: The complex of six MCM proteins is one of several proteins that must be bound
 phosphorylated and dissociate from the chromatin.
C;Genetics:
A;Gene: GDB:MCM4; CDC21; CDC54
A;Cross-references: GDB:433798
C;Complex: The predominant form is a heterohexamer of MCM2, MCM3, MCM4, MCM5, MCM6, and
C;Function:
A;Description: part of the replication licensing system that permits DNA replication to
C;Superfamily: replication licensing factor MCM4; MCM homology
C;Keywords: cell cycle control; DNA replication initiation; heterohexamer; nucleus; phos
F;271-764/Domain: MCM homology <MCM>

```
  Query Match             39.4%;  Score 1814.5;  DB 1;  Length 863;
  Best Local Similarity   44.2%;  Pred. No. 4.9e-94;
  Matches  401;  Conservative  157;  Mismatches  266;  Indels   83;  Gaps   18;

Qy         30 VTSPSFDNAASPV-AGRRAVRQTPTSAVRRR-GRETDSARRRRSRSRSLG---------- 77
              :::||    |::|     | |   | ||      |     | :  | |||    | | |
Db          1 MSSP----ASTPSRRGSRRGRATPAQTPRSEDARSSPSQRRRGEDSTSTGELQPMPTSPG 56

Qy         78 ---------NSVYSSPYDAGTPGTP-GTPVATPV-YATPVGTPMGTP-SFHRGTPQYKQR 125
                       : ::|||     :     |      |::|: | ||      |||  ||||  :||
Db         57 VDLQSTAAQDVLFSSPPQMHSSAIPLDFDVSSPLTYGTPSSRVEGTPRSGVRGTP-VRQR 115

Qy        126 SELGSQGKPLHRRRRSQSREPGHRSPSREPSADGRPSESAEPDDTLGGEYAYVWGTNVNI 185
               :|||  | |     :|                 || :|      :     |:   :|||:||:
Db        116 PDLGSAQKGLQVDLQS----------------DGAAAEDIVASEQSLGQKLVIWGTDVNV 159

Qy        186 PDVLRAIRRFLHNY--------RSSAHDLNSK-YIQIIEETVEREEDTLNIDMSDIYDHD 236
                        :|||   :           :    |:     |:| : |      |   ||::    |    |
Db        160 AACKENFQRFLQRFIDPLAKEEENVGIDITEPLYMQRLGEINVIGEQFLNVNCEHIKSFD 219

Qy        237 PDLYAKIVRYPLDIIPLLDTECQEVATSLLP--TFEKHIEARPFNLKASVHMRELNPSDI 294
              :|| ::: |:  ||  :|||     | |  |     |  ||||      :  :|| ||| ||
Db        220 KNLYRQLISYPQEVIPTFDMAVNEIFFDRYPDSILEHQIQVRPFNALKTKNMRNLNPEDI 279

Qy        295 DKLVSVKGMVIRCSSIIPEIKGAFFKCLVCGHSPPLVTVVKGRVEEPTRCEKPECAARNA 354
              |:|::: ||||| |:|||:: |||:| || |:     | : :||: ||: | :  |     ::
Db        280 DQLITISGMVIRTSQLIPEMQEAFFQCQVCAHT-TRVEMDRGRIAEPSVCGR--CHTTHS 336

Qy        355 MSLIHNRCTFANKQIVRLQETPDAIPEGETPHTVSMCLYNTMVDAVKPGDRIEVTGVFKA 414
              |:||||| |::||:::|||:|: :| |:||||: :  :| :|| |:||||: |||::|
Db        337 MALIHNRSLFSDKQMIKLQESPEDMPAGQTPHTVILFAHNDLVDKVQPGDRVNVTGIYRA 396

Qy        415 MAVRVGPNQRTLRALYKTYIDCVHVKKSDRGRLQTEDPMEMDKENDMYAGYHESDTSEAA 474
               : :|| |     ::::|||:|| :| :|:| ||                  |   |    :
Db        397 VPIRVNPRVSNVKSVYKTHIDVIHYRKTDAKRLH--------------GLDEEAEQKLF 441

Qy        475 NEAKIQKLKELSKLPGIYDRLSRSLAPSIWELEDIKKGLLCQLFGGKAKKI--PSGASFR 532
               :| :::  |||||: | ||:||: :|||||:| ||||||:| |||||  |          ||
Db        442 SEKRVELLKELSRKPDIYERLASALAPSIYEHEDIKKGILLQLFGGTRKDFSHTGRGKFR 501

Qy        533 GDINVLLVGDPGTSKSQLLQYVHKIAPRGIYTSGRGSSAVGLTAYVTKDPETRETVLESG 592
               :||:||  |||||||||||||||: : |||  ||||:|||||||||||| |||||: ||::|
Db        502 AEINILLCGDPGTSKSQLLQYVYNLVPRGQYTSGKGSSAVGLTAYVMKDPETRQLVLQTG 561

Qy        593 ALVLSDRGICCIDEFDKMSDNARSMLHEVMEQQTVSVAKGGIIASLNARTSVLACANPSG 652
              |||||| |||||||||||::: ||:||||||||||:|:|| |||  |||||||||| |||
Db        562 ALVLSDNGICCIDEFDKMNESTRSVLHEVMEQQTLSIAKAGIICQLNARTSVLAAANPIE 621

Qy        653 SRYNARLSVIDNIQLPPTLLSRFDLIYLMLDKPDEQNDRRLARHLVALHYENYEVSKQDA 712
               |::| : : |:||||| ||||||||||:||||  ||  ||||| |||:|:: | ::::
```

```
R;Bowman, S.
submitted to the EMBL Data Library, June 1995
A;Reference number: S57541
A;Accession: S57553
A;Molecule type: DNA
A;Residues: 1-933 <BOW>
A;Cross-references: UNIPARC:UPI0000127256; EMBL:Z49919; NID:g887584; PIDN:CAA90164.1; PI
A;Experimental source: strain AB972
R;Coxon, A.; Maundrell, K.; Kearsey, S.E.
Nucleic Acids Res. 20, 5571-5577, 1992
A;Title: Fission yeast cdc21(+) belongs to a family of proteins involved in an early ste
A;Reference number: S26640; MUID:93087163; PMID:1454522
A;Accession: S26641
A;Molecule type: DNA
A;Residues: 571-646 <COX>
A;Cross-references: UNIPARC:UPI0000168B52; EMBL:Z15032; NID:g3481; PIDN:CAA78750.1; PID:
C;Comment: The complex of six MCM proteins is one of several proteins that must be bound
 phosphorylated and dissociate from the chromatin.
C;Genetics:
A;Gene: SGD:CDC54; HCD21; MIPS:YPR019w
A;Cross-references: SGD:S0006223; MIPS:YPR019w
A;Map position: 16R
C;Complex: The predominant form is a heterohexamer of MCM2 (PIR:S45757), MCM3 (PIR:A3637
component of replication licensing factor.
C;Function:
A;Description: MCM4 is a component of the replication licensing factor that permits DNA
C;Superfamily: replication licensing factor MCM4; MCM homology
C;Keywords: cell cycle control; DNA replication initiation; heterohexamer; nucleus; phos
F;314-828/Domain: MCM homology <MCM>

  Query Match             39.8%;  Score 1833;  DB 1;  Length 933;
  Best Local Similarity   40.9%;  Pred. No. 5e-95;
  Matches  393;  Conservative  178;  Mismatches  281;  Indels  110;  Gaps   14;

Qy          2 ENNDALDIGAVSSPYPSQSEGVSTPLPQVTSPSFDNAASPVAG----------------- 44
              |:|::        |||      :  |     ||::||:   ::|    |
Db         11 EDNNS------SSPVVPNPDSVP---PQLSSPALFYSSSSSQGDIYGRNNSQNLSQGEGN 61

Qy         45 -RRAVRQTP---TSAVRRRGRETDSARRRRSRSRSLGNSVYSSPYDAGTPGTPGTPVATP 100
               | |:  :|      |: :|:  :    :: |: | ||  ::    | | :          | ::
Db         62 IRAAIGSSPLNFPSSSQRQNSDVFQSQGRQGRIRSSASASGRSRYHSDLRSDRALPTSS- 120

Qy        101 VYATPVGTPMGTPSFHRGTPQYKQRSELGSQGKPLHRRRRSQSREPGHRSPSREPSADGR 160
                                          | ||  |:        ||:          || |      | |
Db        121 -------------------------SSLGRNGQNRVHMRRNDIHTSDLSSPRRIVDFDTR 155

Qy        161 --------------PSESAEPDDTLGGEYAYVWGTNVNIPDVLRAIRRFLHNYR------ 200
                            |||::||            :|||||:|  :       | || :::
Db        156 SGVNTLDTSSSSAPPSEASEP-------LRIIWGTNVSIQECTTNFRNFLMSFKYKFRKI 208

Qy        201 --------SSAHDLNSKYIQIIEETVEREEDTLNIDMSDI--YDHDPDLYAKIVRYPLDI 250
                       ::   |        ||: :  |  |       ||:  ::    |   ||| ::: || ::
Db        209 LDEREEFINNTTDEELYYIKQLNEMRELGTSNLNLDARNLLAYKQTEDLYHQLLNYPQEV 268

Qy        251 IPLLDTECQEVATSL---------LPTFE-KHIEARPFNLKASVHMRELNPSDIDKLVSV 300
               | ::|    ::   ||            |     |  |   :  ||:|: :      ||||||:||||||:::
Db        269 ISIMDQTIKDCMVSLIVDNNLDYDLDEIETKFYKVRPYNVGSCKGMRELNPNDIDKLINL 328

Qy        301 KGMVIRCSSIIPEIKGAFFKCLVCGHSPPLVTVVKGRVEEPTRCEKPECAARNAMSLIHN 360
              ||:|:| : :||::| ||||| || |:       | : :| ::|| |||: :|     |:||||||
Db        329 KGLVLRSTPVIPDMKVAFFKCNVCDHT-MAVEIDRGVIQEPARCERIDCNEPNSMSLIHN 387

Qy        361 RCTFANKQIVRLQETPDAIPEGETPHTVSMCLYNTMVDAVKPGDRIEVTGVFKAMAVRVG 420
              ||:||:||:::||||| :|:|:|||::|:|:|:|: :||: : ||||||| |::: :|
Db        388 RCSFADKQVIKLQETPDFVPDGQTPHSISLCVYDELVDSCRAGDRIEVTGTFRSIPIRAN 447

Qy        421 PNQRTLRALYKTYIDCVHVKKSDRGRLQTEDPMEMDKENDMYAGYHESDTSEAANEAKIQ 480
                 || |::||||:| |||||       ||    :             :       ::|  :       :   :
Db        448 SRQRVLKSLYKTYVDVVHVKKVSDKRLDVDTSTIEQELMQNKVDHNEVEEVRQITDQDLA 507

Qy        481 KLKELSKLPGIYDRLSRSLAPSIWELEDIKKGLLCQLFGGKAKKIPSGASFRGDINVLLV 540
              |:|:|:      :|   |:|:||||:||||:|||:| |||||     |       |    :||||:||
Db        508 KIREVAAREDLYSLLARSIAPSIYELEDVKKGILLQLFGGTNKTFTKGGRYRGDINILLC 567

Qy        541 GDPGTSKSQLLQYVHKIAPRGIYTSGRGSSAVGLTAYVTKDPETRETVLESGALVLSDRG 600
              ||| |||||:|||||||  |||:||||:||||||||||||:|:|  :|::  |||||||||||  |
Db        568 GDPSTSKSQILQYVHKITPRGVYTSGKGSSAVGLTAYITRDVDTKQLVLESGALVLSDGG 627

Qy        601 ICCIDEFDKMSDNARSMLHEVMEQQTVSVAKGGIIASLNARTSVLACANPSGSRYNARLS 660
              :|||||||||||: ||:||||||||||:|:|:|| |||  :||||:|:|| ||| |||||   |
Db        628 VCCIDEFDKMSDSTRSVLHEVMEQQTISIAKAGIITTLNARSSILASANPIGSRYNPNLP 687

Qy        661 VIDNIQLPPTLLSRFDLIYLMLDKPDEQNDRRLARHLVALHYEN--YEVSKQDALDLQTL 718
              | :|| ||| ||||||:||:|||  ||:|||  ||:||   |: |:     :|: | | :: |
Db        688 VTENIDLPPPLLSRFDLVYLVLDKVDEKNDRELAKHLTNLYLEDKPEHISQDDVLPVEFL 747

Qy        719 TAYITYARQHVHPTLSDEAAEDLINGYVEMRQKGNFPGSSKKVITATPRQLESMIRISEA 778
              | ||:||::|:|| ::: |   :|:   || ||: |:   | :| |||| ||||||||::||
Db        748 TMYISYAKEHIHPIITEAAKTELVRAYVGMRKMGDDSRSDEKRITATTRQLESMIRLAEA 807

Qy        779 LARMRFSEVVEKVDAAEAVRLLDVALQQSATDHATGTIDMDLITTGVSASERIRRANLLA 838
              |:|:   |||  |  |||||:  |::  |||  |||  ||||:|:  || |   :|  : :|
Db        808 HAKMKLKNVVELEDVQEAVRLIRSAIKDYATDPKTGKIDMNLVQTGKSVIQRKLQEDLSR 867

Qy        839 ALRELIADKISPGSSSGLKTSQLLEDIRSQSSVDVSLQDIKNALGSLQGEGFLTVHGDIV 898
               :  ::|  |:    :|    :   ::|::  |     |     |    ||: ||   ||  |   : | |: |
Db        868 EIMNVLKDQ----ASDSMSFNELIKQINEHSQDRVESSDIQEALSRLQQEDKVIVLGEGV 923

Qy        899 KR 900
               :|
Db        924 RR 925
```

```
RESULT 4
S56766
replication licensing factor MCM4 - mouse
N;Alternate names: cell division cycle control protein CDC21/CDC54
C;Species: Mus musculus (house mouse)
C;Date: 13-Aug-1999 #sequence_revision 13-Aug-1999 #text_change 09-Jul-2004
C;Accession: S56766
R;Kimura, H.; Takizawa, N.; Nozaki, N.; Sugimoto, K.
Nucleic Acids Res. 23, 2097-2104, 1995
A;Title: Molecular cloning of cDNA encoding mouse Cdc21 and CDC46 homologs and characte
A;Reference number: S56766; MUID:95334361; PMID:7610039
A;Accession: S56766
A;Status: nucleic acid sequence not shown; translation not shown
A;Molecule type: mRNA
A;Residues: 1-862 <KIM>
A;Cross-references: UNIPROT:P49717; UNIPARC:UPI00000041B9; EMBL:D26089; NID:g940405; PII
A;Note: the nucleotide sequence was submitted to the EMBL Data Library, December 1993
C;Comment: The complex of six MCM proteins is one of several proteins that must be bounc
 phosphorylated and dissociate from the chromatin.
C;Complex: The predominant form is a heterohexamer of MCM2, MCM3, MCM4, MCM5, MCM6, and
C;Function:
A;Description: part of the replication licensing system that permits DNA replication to
C;Superfamily: replication licensing factor MCM4; MCM homology
C;Keywords: cell cycle control; DNA replication initiation; heterohexamer; nucleus; phos
F;270-763/Domain: MCM homology <MCM>

  Query Match             39.5%;  Score 1821;  DB 1;  Length 862;
  Best Local Similarity   43.7%;  Pred. No. 2.1e-94;
  Matches  401;  Conservative  154;  Mismatches  258;  Indels  104;  Gaps   18;

Qy         30 VTSPSFDNAASPVAGRRAVRQ---TPTSAVRRRGRETDSARRRRSRSRSLG--------- 77
              ::||     |::|    ||: |:     ||| ::|         :     ||||      | |
Db          1 MSSP----ASTP--SRRSSRRGRVTPTQSLRSEESRSSPNRRRRGEDSSTGELLPMPTSP 54

Qy         78 ----------NSVYSSP-----------YDAGTPGTPGTPVATPVYATPVGTPMGTP-SF 115
                          |:::|||                :|   :| |         | ||        ||| |
Db         55 GADLQSPPAQNALFSSPPQMHSLAIPLDFDVSSPLT---------YGTPSSRVEGTPRSG 105

Qy        116 HRGTPQYKQRSELGSQGKPLHRRRRSQSREPGHRSPSREPSADGRPSESAEPDDTLGGEY 175
                ||||  :|| :|||  | |       :|                ||   :|   |  :    |:
Db        106 VRGTP-VRQRPDLGSARKGLQVDLQS-----------------DGAAAEDIVPSEQSLGQK 148

Qy        176 AYVWGTNVNIPDVLRAIRRFLHNY--------RSSAHDLNSK-YIQIIEETVEREEDTLN 226
```

```
Qy        569 SSAVGLTAYVTKDPETRETVLESGALVLSDRGICCIDEFDKMSDNARSMLHEVMEQQTVS 628

Db        409 SSAVGLTAYVAKDPETGETVLESGALVLSDRGICCIDEFDKMSDSARSMLHEVMEQQTVS 468

Qy        629 VAKGGIIASLNARTSVLACANPSGSRYNARLSVIDNIQLPPTLLSRFDLIYLMLDKPDEQ 688

Db        469 IAKAGIIASLNARTSVLACANPSGSRYNPRLSVIENIHLPPTLLSRFDLIYLILDKPDEQ 528

Qy        689 NDRRLARHLVALHYENYEVSKQDALDLQTLTAYITYARQHVHPTLSDEAAEDLINGYVEM 748

Db        529 TDRRLAKHIVALHFENAESAQEEAIDITTLTTYVSYARKNIHPKLSDEAAEELTRGYVEL 588

Qy        749 RQKGNFPGSSKKVITATPRQLESMIRISEALARMRFSEVVEKVDAAEAVRLLDVALQQSA 808

Db        589 RKAGKFAGSSKKVITATPRQIESLIRLSEALARMRFSEWVEKHDVDEAFRLLRVAMQQSA 648

Qy        809 TDHATGTIDMDLITTGVSASERIRRANLLAALRELIADKISPGSSSGLKTSQLLEDIRSQ 868

Db        649 TDHATGTIDMDLINTGVSASERMRRDTFASSIRDIALEKI----TMEMLLLQLLEELKKH 704

Qy        869 S---SVDVSLQDIKN 880

Db        705 GGNINTEIHLHDVSH 719
```

RESULT 2
S26640
replication licensing factor MCM4 - fission yeast (Schizosaccharomyces pombe)
N;Alternate names: cell division cycle control protein CDC21/CDC54
C;Species: Schizosaccharomyces pombe
C;Date: 13-Aug-1999 #sequence_revision 13-Aug-1999 #text_change 09-Jul-2004
C;Accession: S26640; T41090; S21829
R;Coxon, A.; Maundrell, K.; Kearsey, S.E.
Nucleic Acids Res. 20, 5571-5577, 1992
A;Title: Fission yeast cdc21(+) belongs to a family of proteins involved in an early ste
A;Reference number: S26640; MUID:93087163; PMID:1454522
A;Accession: S26640
A;Molecule type: DNA
A;Residues: 1-909 <COX>
A;Cross-references: UNIPROT:P29458; UNIPARC:UPI000017332F; EMBL:X58824; NID:g4926; PID:g
R;McDougall, R.C.; Rajandream, M.A.; Barrell, B.G.; Zimmermann, W.; Wambutt, R.
submitted to the EMBL Data Library, August 1999
A;Reference number: Z21822
A;Accession: T41090
A;Status: translated from GB/EMBL/DDBJ
A;Molecule type: DNA
A;Residues: 1-546 <MCD>
A;Cross-references: UNIPARC:UPI0000161F1A; EMBL:AL109957; PIDN:CAB53089.1; GSPDB:GN00068
A;Experimental source: strain 972h(-); cosmid c16A11
C;Comment: The complex of six MCM proteins is one of several proteins that must be bound
 phosphorylated and dissociate from the chromatin.
C;Genetics:
A;Gene: cdc21; SPDB:SPCC16A11.17
A;Map position: 3
C;Complex: predominant form is a heterohexamer of MCM2, MCM3, MCM4, MCM5, MCM6, and MCM7
C;Function:
A;Description: part of the replication licensing system that permits DNA replication to
C;Superfamily: replication licensing factor MCM4; MCM homology
C;Keywords: cell cycle control; DNA replication initiation; heterohexamer; nucleus; phos
F;291-805/Domain: MCM homology <MCM>

```
  Query Match            40.8%;  Score 1881;  DB 1;  Length 909;
  Best Local Similarity  44.0%;  Pred. No. 9.7e-98;
  Matches  411;  Conservative  164;  Mismatches  241;  Indels  118;  Gaps   21;

Qy         38 AASPVAGRRAVRQTPTSAVRRRGRETDSARRRRSRSRSLGNSVYSSP--YDAGTPG---- 91

Db          2 SSSQQSGRANELRTP-------GRANSSSR----------EAVDSSPLFFPASSPGSTRL 44

Qy         92 -TPGTPVATPVYATPVGTPMGTP---------------------SFHRGTPQYKQRSEL 128

Db         45 TTPRTTARTPLASSPLLFESSSPGPNIPQSSRSHLLSQRNDLFLDSSSQRTPRSTRRGDI 104

Qy        129 GS--QGKPLHRRRRSQSREPGHRSPSR--EPSADGRPSESAEP-----DDTLGGEYAYVW 179

Db        105 HSSVQMSTPSRRREVDPQRPGVSTPSSLLFSGSDALTFSQAHPSSEVADDTV----RVIW 160

Qy        180 GTNVNIPDVLRAIRRFLHNYRS--------------SAHDLNSKYIQIIEETVEREEDTL 225

Db        161 GTNVSIQESIASFRGFLRGFKKKYRPEYRNELMPPPDAEQL--VYIEALRNMRIMGLEIL 218

Qy        226 NIDMSDIYDHDP--DLYAKIVRYPLDIIPLLDTECQEVATSLLPTFE----------KHI 273

Db        219 NLDVQDLKHYPPTKKLYHQLYSYPQEIIPIMDQTIKDVMLDLLGTNPPEDVLNDIELKIY 278

Qy        274 EARPFNLKASVHMRELNPSDIDKLVSVKGMVIRCSSIIPEIKGAFFKCLVCGHSPPLVTV 333

Db        279 KIRPFNLEKCINMRDLNPGDIDKLISIKGLVLRCTPVIPDMKQAFFRCSVCGHC-VTVEI 337

Qy        334 VKGRVEEPTRCEKPECAARNAMSLIHNRCTFANKQIVRLQETPDAIPEGETPHTVSMCLY 393

Db        338 DRGRIAEPIKCPREVCGATNAMQLIHNRSEFADKQVIKLQETPDVVPDGQTPHSVSLCVY 397

Qy        394 NTMVDAVKPGDRIEVTGVFKAMAVRVGPNQRTLRALYKTYIDCVHVKKSDRGRLQTEDPM 453

Db        398 DELVDSARAGDRIEVTGIFRCVPVRLNPRMRTVKSLFKTYVDVVHIKKQDKRRLGT-DPS 456

Qy        454 EMDKENDMYAGYHESDTSEAA----------NEAKIQKLKELSKLPGIYDRLSRSLAPSI 503

Db        457 TL-----------ESDIAEDAALQIDEVRKISDEEVEKIQQVSKRDDIYDILSRSLAPSI 505

Qy        504 WELEDIKKGLLCQLFGGKAKKIPSGAS--FRGDINVLLVGDPGTSKSQLLQYVHKIAPRG 561

Db        506 YEMDDVKKGLLLQLFGGTNKSFHKGASPRYRGDINILMCGDPSTSKSQILKYVHKIAPRG 565

Qy        562 IYTSGRGSSAVGLTAYVTKDPETRETVLESGALVLSDRGICCIDEFDKMSDNARSMLHEV 621

Db        566 VYTSGKGSSAVGLTAYITRDQDTKQLVLESGALVLSDGGICCIDEFDKMSDATRSILHEV 625

Qy        622 MEQQTVSVAKGGIIASLNARTSVLACANPSGSRYNARLSVIDNIQLPPTLLSRFDLIYLM 681

Db        626 MEQQTVTVAKAGIITTLNARTSILASANPIGSKYNPDLPVTKNIDLPPTLLSRFDLVYLI 685

Qy        682 LDKPDEQNDRRLARHLVALHYEN--YEVSKQDALDLQTLTAYITYARQHVHPTLSDEAAE 739

Db        686 LDRVDETLDRKLANHIVSMYMEDTPEHATDMEVFSVEFLTSYITYARNNINPVISEEAAK 745

Qy        740 DLINGYVEMRQKGNFPGSSKKVITATPRQLESMIRISEALARMRFSEVVEKVDAAEAVRL 799

Db        746 ELVNAYVGMRKLGEDVRASEKRITATTRQLESMIRLSEAHAKMHLRNVVEVGDVLEAARL 805

Qy        800 LDVALQQSATDHATGTIDMDLITTGVSASERIRRANLLAALRELIADKISPGSSSGLKTS 859

Db        806 IKTAIKDYATDPATGKISLDLIY--VNERETLVPEDMVKELANLISNLTVGGKT--MLVS 861

Qy        860 QLLEDIRSQSSVDVSLQDIKNALGSLQGEGFLTV 893

Db        862 QLLTRFREQSSTRLDASDFEACLGALERRGRIKV 895
```

RESULT 3
S56050
replication licensing factor MCM4 [validated] - yeast (Saccharomyces cerevisiae)
N;Alternate names: cell division control protein CDC21/CDC54; protein YP9531.13; protein
C;Species: Saccharomyces cerevisiae
C;Date: 13-Aug-1999 #sequence_revision 13-Aug-1999 #text_change 09-Jul-2004
C;Accession: S56050; S57553; S26641; S25527
R;Dalton, S.
submitted to the EMBL Data Library, September 1994
A;Description: Cdc47 and Cdc54 belong to a family of proteins essential for initiation o
A;Reference number: S56049
A;Accession: S56050
A;Molecule type: DNA
A;Residues: 1-933 <DAL>
A;Cross-references: UNIPROT:P30665; UNIPARC:UPI0000127256; EMBL:U14731; NID:g608170; PID

OM protein - protein search, using sw model

```
Run on:         December  7, 2005, 17:28:25 ; Search time 45 Seconds
                                          (without alignments)
                                          1926.473 Million cell updates/sec

Title:          US-10-768-511-8
Perfect score:  4609
Sequence:       1 MENNDALDIGAVSSPYPSQS..........LGSLQGEGFLTVHGDIVKRV 901

Scoring table:  BLOSUM62
                Gapop 10.0 , Gapext 0.5

Searched:       283416 seqs, 96216763 residues

Total number of hits satisfying chosen parameters:      283416

Minimum DB seq length: 0
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%
                 Maximum Match 100%
                 Listing first 45 summaries

Database :       PIR_80:*
                 1:  pir1:*
                 2:  pir2:*
                 3:  pir3:*
                 4:  pir4:*
```

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	% Query Match	Length	DB	ID	Description
1	2332	50.6	720	2	C84540	probable CDC21 pro
2	1881	40.8	909	1	S26640	replication licens
3	1833	39.8	933	1	S56050	replication licens
4	1821	39.5	862	1	S56766	replication licens
5	1814.5	39.4	863	1	S65954	replication licens
6	1801.5	39.1	863	1	S64720	replication licens
7	1784.5	38.7	858	2	T47223	replication licens
8	1689.5	36.7	866	1	S59872	replication licens
9	1064	23.1	407	2	T50408	cdc21 protein [imp
10	1039	22.5	686	2	H90226	minichromosome mai
11	1020.5	22.1	666	1	D69103	DNA helicase (EC 3
12	1016.5	22.1	760	2	T39991	minichromosome mai
13	973	21.1	699	2	B72775	probable DNA repli
14	969	21.0	720	2	T47221	replication licens
15	966	21.0	719	1	JC4580	replication licens
16	963	20.9	775	1	A39631	replication licens
17	961.5	20.9	719	1	S70583	replication licens
18	961.5	20.9	881	2	T26498	hypothetical prote
19	954.5	20.7	720	1	A48723	replication licens
20	947.5	20.6	716	1	T01507	replication licens
21	947.5	20.6	735	1	PC4225	replication licens
22	940	20.4	936	2	E96508	hypothetical prote
23	938	20.4	586	1	E69314	replication licens
24	932	20.2	1017	1	S64219	replication licens
25	930	20.2	904	1	T10067	replication licens
26	928	20.1	734	1	I38080	replication licens
27	927.5	20.1	733	1	S56767	replication licens
28	927.5	20.1	759	2	T24130	hypothetical prote
29	924	20.0	868	1	S45757	replication licens
30	923	20.0	796	2	T47222	replication licens
31	920	20.0	810	2	H88565	protein ZK632.1 [i
32	914.5	19.8	892	2	T50339	mis5 protein [impo
33	910.5	19.8	868	2	T43423	probable replicati
34	909	19.7	845	1	S34027	replication licens
35	903.5	19.6	886	1	JC5085	replication licens
36	901.5	19.6	727	2	G84487	probable DNA repli
37	901.5	19.6	879	2	T41059	replication licens
38	897	19.5	807	1	I51685	replication licens
39	895.5	19.4	730	1	T03920	replication licens
40	882	19.1	830	1	B48723	replication licens
41	879.5	19.1	776	2	T52118	probable replicati
42	871.5	18.9	744	1	I51022	replication licens
43	858.5	18.6	971	1	A36376	replication licens
44	843.5	18.3	795	1	S22804	replication licens
45	842	18.3	812	2	T19446	hypothetical prote

ALIGNMENTS

```
RESULT 1
C84540
probable CDC21 protein [imported] - Arabidopsis thaliana
C;Species: Arabidopsis thaliana (mouse-ear cress)
C;Date: 02-Feb-2001 #sequence_revision 02-Feb-2001 #text_change 09-Jul-2004
C;Accession: C84540
R;Lin, X.; Kaul, S.; Rounsley, S.D.; Shea, T.P.; Benito, M.I.; Town, C.D.; Fujii, C.Y.;
M.; Koo, H.; Moffat, K.S.; Cronin, L.A.; Shen, M.; VanAken, S.E.; Umayam, L.; Tallon, L
euss, D.; Nierman, W.C.; White, O.; Eisen, J.A.; Salzberg, S.L.; Fraser, C.M.; Venter,
Nature 402, 761-768, 1999
A;Title: Sequence and analysis of chromosome 2 of the plant Arabidopsis thaliana.
A;Reference number: A84420; MUID:20083487; PMID:10617197
A;Accession: C84540
A;Status: preliminary
A;Molecule type: DNA
A;Residues: 1-720 <STO>
A;Cross-references: UNIPROT:Q9SIV8; UNIPARC:UPI00000A085A; GB:AE002093; NID:g4544386; P
C;Genetics:
A;Gene: At2g16440
A;Map position: 2

  Query Match             50.6%;  Score 2332;  DB 2;  Length 720;
  Best Local Similarity   61.0%;  Pred. No. 3.3e-123;
  Matches  448;  Conservative  124;  Mismatches  135;  Indels   28;  Gaps    6;

Qy        154 EPSADGRPSESAEPDDTLGGEYAYVWGTNVNIPDVLRAIRRFLHNY---RSSAHDL--NS 208
Db          5 EPLPSSDDGEEDGGDDTTP---TFVWGTNISVQDVKSAIEMFVKHFREARENSDDLFREG 61

Qy        209 KYIQIIEETVEREEDTLNIDMSDIYDHDPDLYAKIVRYPLDIIPLLDTECQEVATSLLPT 268
Db         62 KYMVSIRKVIEIEGEWIDVDAFDVFDYDPDLYNKMVRYPLEVLAIFDIVLMDIVSTINRL 121

Qy        269 FEKHIEARPFNLKASVHMRELNPSDIDKLVSVKGMVIRCSSIIPEIKGAFFKCLVCGHSP 328
Db        122 FEKHVQVRIFNLRTSTSMRNLNPSDIEKMISLKGMIIRSSSIIPEIREAVFRCLVCGYFS 181

Qy        329 PLVTVVKGRVEEPTRCEKPECAARNAMSLIHNRCTFANKQIVRLQETPDAIPEGETPHTV 388
Db        182 DPIIVDRGKISEPPTCLKQECMTKNSMTLVHNRCRFADKQIVRLQETPDEIPEGGTPHTV 241

Qy        389 SMCLYNTMVDAVKPGDRIEVTGVFKAMAVRVGPNQRTLRALYKTYIDCVHVKKSDRGRLQ 448
Db        242 SLLLHDKLVDNGKPGDRIEVTGIYRAMTVRVGPAHRTVKSVFKTYIDCLHIKKASKLRMS 301

Qy        449 TEDPMEMDKENDMYAGYHESDTSEAANEAKIQKLKELSKLPGIYDRLSRSLAPSIWELED 508
Db        302 AEDPMDV-------------DNSLRRVDEDLRKFQELSKQPDIYERLSRSLAPNIWELDD 348

Qy        509 IKKGLLCQLFGGKAKKIPSGASFRGDINVLLVGDPGTSKSQLLQYVHKIAPRGIYTSGRG 568
Db        349 VKKGLLCQLFGGNALNLASGANFRGDINILLVGDPGTSKSQLLQYIHKLSPRGIYTSGRG 408
```

```
;  CURRENT APPLICATION NUMBER: US/10/957,569
;  CURRENT FILING DATE:  2004-09-30
;  PRIOR APPLICATION NUMBER: US 10/950,321
;  PRIOR FILING DATE: 2004-09-23
;  NUMBER OF SEQ ID NOS: 64
;  SOFTWARE: PatentIn version 3.3
; SEQ ID NO 31
;   LENGTH: 824
;   TYPE: PRT
;   ORGANISM: Arabidopsis thaliana
US-10-957-569-31

  Query Match            2.6%;  Score 122;  DB 6;  Length 824;
  Best Local Similarity  20.7%;  Pred. No. 0.36;
  Matches  156;  Conservative   92;  Mismatches  267;  Indels  240;  Gaps   37;

Qy        123 KQRSELGSQGKPLHRRRRSQSREPGH-RSPSREPSADGRPSESAEPDDTLGGEYAYVWGT 181
              |  |  ||  |   ||  | :    :   | | : ||  :  | |   |: |
Db        198 KSSSMEGSSNPPKLRRNSSAAANISNLASSSNQVSAPLKRSSSWSFDEKL---------- 247

Qy        182 NVNIPDVLRAIRRFLHNYRSSAHDLNSKYIQIIEETVEREEDTLNIDMSDIYDHDPDLYA 241
                    ::::: : |  | | |: :   |:: :|  : | : | |           |:
Db        248 ------LVQSLYKVL-AYVSKANPI-VLYLRDVENFLFRSQRTYN------------LFQ 287

Qy        242 KIVRY---PL-----DIIPLLDTECQEVATSLLPTFEKHIEARPFNLKASVHMRELNPSD 293
              |:::    |:      |: |   : ||:   |   |  :|: ||             ||
Db        288 KLLQKLSGPVLILGSRIVDLSSEDAQEIDEKLSAVFPYNIDIRP-------------PED 334

Qy        294 IDKLVSVKGMVIRCSSIIP---------EIKGAFFKCLVCGHSPPL----VTVVKGRVEE 340
                 ||| |  : |  ::|          |:      |:|     :       |:   :||
Db        335 ETHLVSWKSQLERDMNMIQTQDNRNHIMEVLSE--NDLICDDLESISFEDTKVLSNYIEE 392

Qy        341 PT---------RCEKPECAARN------AMSLIHNRCTFAN---------KQIVRLQETP 376
                       : ||   ||        ::||  |    |           ||  : : :
Db        393 IVVSALSYHLMNNKDPE--YRNGKLVISSISLSHGFSLFREGKAGGREKLKQKTKEESSK 450

Qy        377 DAIPEGETPH--TVSMCLYNTMVDAVKPGDRIEVTGVFKAMAVRVGPNQRTLRALYKTYI 434
              :    |   |   | |:   ::  :  |        :||    | | |:    : :    |
Db        451 EVKAESIKPETKTESVTTVSSKEEPEKAKAEKVT----PKAPEVAPDNEFEKRIRPEVI 506

Qy        435 DC--VHVKKSDRGRLQTEDPMEMDKENDMYAGYHESDTSEAANEAKIQKLKELSKLPGIY 492
                 ::|   | |                           | :| | : |:|| ||
Db        507 PAEEINVTFKDIG--------------------------ALDEIK-ESLQELVMLP--- 535

Qy        493 DRLSRSLAPSIWELEDIKKGLLCQLFGGKAKKIPSGASFRGDINVLLVGDPGTSKSQLLQ 552
                 | |   | ::               || |      ||   :|| | ||| |: | :
Db        536 --LRR---PDLFT------------GGLLKPC------RG---ILLFGPPGTGKTMLAK 568

Qy        553 YVHKIAPRGIYTSGRGSSAVGLTAYVTK---DPETRETVLESGALVLSDRGICCIDEFDK 609
               : | |        |     | ::   :|   : |     | : |  :|   |  :|| |
Db        569 AIAKEA-------GASFINVSMSTITSKWFGEDEKNVRALFTLASKVSPT-IIFVDEVDS 620

Qy        610 M-SDNARSMLHEVMEQQTVSVAK--GGIIASLNARTSVLACANPSGSRYNARLSVIDNIQ 666
              |     |   || | :          |::    |  ||| |
Db        621 MLGQRTRVGEHEAMRKIKNEFMSHWDGLMTKPGERILVLAATNRP-------------FD 667

Qy        667 LPPTLLSRFDLIYLMLDKPDEQNDRRLARHLVA-------LHYENYEVSKQ-----DALD 714
              |   :: ||:  :|:  |  :|  :: | |:|         | |:   :  :      |  :
Db        668 LDEAIIRRFER-RIMVGLPAVENREKILRTLLAKEKVDENLDYKELAMMTEGYTGSDLKN 726

Qy        715 LQTLTAY------------ITYARQHVHPTLSDEAAEDLINGYVEMRQKGNFPGSSKKV 761
              | |  ||             |  ::   || : |  |                 |  ::|
Db        727 LCTTAAYRPVRELIQQERIKDTEKKKQREPTKAGEEDE----------------GKEERV 770

Qy        762 ITATPRQLESMIRISEALARMRFSEVVEKVDAAEA 796
              ||  |            | |  |  |   :| |: |
Db        771 ITLRP------------LNRQDFKEAKNQVAASFA 793

Search completed: December  7, 2005, 17:47:50
Job time : 17 secs
```

```
Qy        423 QR--------TLRALYKTYIDCVHVKKSDRGRLQTEDPMEM--DKENDM-------YAGY 465
Db        871 LKRGIAGVLARIGALLRRVLQRLAGKAG--GAMSDADVFAMLADLHGNVEGARDAPWGGN 928
Qy        466 HES-----DTSEAANEAKIQKLKELSKLPGIYDRLSRSLAPSIWELEDIKKGLLCQLFGG 520
Db        929 HRAVMFARAEDGAAERSKSESLEKLRRAETI--RISGREVPEGGNLREYKRNAL------ 980
Qy        521 KAKKIPSGASFRGDINVLLVGDPGTSKSQLLQYVHKIAPRGIYTSGRGSSAVGLTAYVTK 580
Db        981 -----EYGKSLRG------------------PYVNKDTGREI-SLGRS----GITEILRH 1012
Qy        581 DPETRETVLESGALVLSDRGICCIDEFDKMSDNARSMLHEVMEQQTVSVAKGGIIASLNA 640
Db       1013 DYKDAEHLQSIAAIPQIIENAVYIDTLPN-EDLAKNGDIQGYEYYVSGLNVGG--ADYTV 1069
Qy        641 RTSVLACANPSGSRY----------NARLSVIDNIQLPPTLLSRFDLIYLMLDKPDEQND 690
Db       1070 RAAVAVSRN--GNRYYDHKLTKIEKGNLLSLLDRVSTTGASESKSPLSGI--------DD 1119
Qy        691 RRLARHLV--------ALHYENYEVSKQDALDLQTLTAYITYARQHVHPTLSD------- 735
Db       1120 KRLLQILQDKDAGKGGIADFDTEAVRFSRAANIEAAIGRITGKKSDLRNALKDRWDASKG 1179
Qy        736 --------EAAEDLINGYVE-MRQKGN----FPGSSKKVITATPRQLESMIRISE----A 778
Db       1180 IQLQFLGRRQIEDIYGGVLDGLKEYGRLSELFGADANKAVTEADKVVREWGRLKEEDAKA 1239
Qy        779 LARMRFSEVVEKVDA-----AEAVRLLDVALQQSATDHATGTIDMDLITTGVSASERIRR 833
Db       1240 LADLMHDATLAKVDADPLMRKDAQKRLDGI--RTALDIADGKIEKAEAAV-ASAGARIAR 1296
Qy        834 ANLLAALRELIADK 847
Db       1297 ADAAYNKAQRAADK 1310
```

RESULT 14
US-10-821-234-1145
; Sequence 1145, Application US/10821234
; Publication No. US20050255114A1
; GENERAL INFORMATION:
; APPLICANT: Labat, Ivan
; APPLICANT: Stache-Crain, Birgit
; APPLICANT: Andarmani, Susan
; APPLICANT: Tang, Y. Tom
; TITLE OF INVENTION: Methods for Diagnosis and Treatment of Preeclampsia
; FILE REFERENCE: 821A
; CURRENT APPLICATION NUMBER: US/10/821,234
; CURRENT FILING DATE: 2004-04-07
; PRIOR APPLICATION NUMBER: US 60/462,047
; PRIOR FILING DATE: 2003-04-07
; NUMBER OF SEQ ID NOS: 1704
; SOFTWARE: pt_SEQ_genes Version 1.0
; SEQ ID NO 1145
; LENGTH: 1299
; TYPE: PRT
; ORGANISM: Homo sapiens
US-10-821-234-1145

```
  Query Match             2.7%;  Score 125.5;  DB 6;  Length 1299;
  Best Local Similarity   18.3%;  Pred. No. 0.4;
  Matches  190;  Conservative  144;  Mismatches  340;  Indels  365;  Gaps   52;

Qy         10 GAVSSPYPSQSEGVSTPLPQVTSPSFDNAASPVAGRRAVRQTPTSAVRRRGRETDSARRR 69
Db         16 GTQPSTARRREEGPPPPSPD--GASSDAEPEPPSGR---TESPATAAETASEELD----- 65
Qy         70 RSRSRSLGNSVYSSPYDAGTPGTPGTPVATPVYATPVGTP--MGTPSFHRGTPQYKQRSE 127
Db         66 ---NRSLEEILNSIP-----------PPPPPAMTNEAGAPRLMITHIVNQNFKSYAGEKI 111
Qy        128 LGSQGKPLHRRRRSQSREPGHRSPSREPSADGRPSESAEPDDTLGGEYAYVWGTN----V 183
Db        112 LG----PFHKR-----------------------------------FSCIIGPNGSGKS 131
Qy        184 NIPDVLRAIRRFLHNYRSSAHDLNSKYIQIIEETVEREED----TLNIDMSDIYDHDPDL 239
Db        132 NVIDSM----LFVFGYR--AQKIRSKKLSVLIHNSDEHKDIQSCTVEVHFQKIIDKEGDD 185
Qy        240 YAKIVRYPLDIIP-----LLDTECQEVATSLLPTFEKHIEARPFNLKASVHMRELNPSDI 294
Db        186 Y--------EVIPNSNFYVSRTACRD------NTSVYHISGKKKTFKDVGNLLRSHGIDL 231
Qy        295 D--KLVSVKGMVIRCSSIIPEIKGAFFKCLVCGHSPPLVTVVK-----GRVEEPTR--CE 345
Db        232 DHNRFLILQGEVEQIAMMKP--KGQ------TEHDEGMLEYLEDIIGCGRLNEPIKVLCR 283
Qy        346 KPECAARNAMSLIHNRCTFANKQIVRLQETPDAIPEGETPHTVS--------------MC 391
Db        284 RVEI-------LNEHRGDKLNR-VKMVEKEKDAL-EGEKNIAIEFLTLENEIFRKKNHVC 334
Qy        392 LY------NTMVDAVKPGDRIE-----------VTGVFKAMAVRVGPNQRTLRALYK-- 431
Db        335 QYYIYELQKRIAEMETQKEKIHEDTKEINEKSNILSNEMKAKNKDVKDTEKKLNKITKFI 394
Qy        432 -------TYIDCVHVKKSDRGRLQT------EDPMEMDKE-----------NDMYAGYH 466
Db        395 EENKEKFTQLDLEDVQVREKLKHATSKAKKLEKQLQKDKEKVEEFKSIPAKSNNII---- 450
Qy        467 ESDTSEAANEAKIQKLKELSKLPGIYDRLSRSLAPSIWELEDIKKGLLCQLFGGKAKKIP 526
Db        451 -NETTTRNNALEKEKEKEEKKLKEVMDSLKQETQGLQKEKESREKELM-----GFSKSVN 504
Qy        527 SGASFRGDINVLLVGDPGTSKSQLLQYVHKIAPRGIYTSGRGSSAVGLTAYVTKDPE--- 583
Db        505 EARS-KMDV----------AQSEL----------DIYLS-RHNTAV---SQLTKAKEALI 539
Qy        584 -TRETVLESGALVLSDRGICC------------IDEFDKMSDNARSMLHEV---MEQQTV 627
Db        540 AASETLKERKAAIRDIEGKLPQTEQELKEKEKELQKLTQEETNFKSLVHDLFQKVEEAKS 599
Qy        628 SVA---------------KGGIIASLNARTSVLACANPSGSRYNARLS---------VI 662
Db        600 SLAMNRSRGKVLDAIIQEKKSGRIPGIYGRLGDLGAID---EKYDVAISSCCHALDYIVV 656
Qy        663 DNIQLPP---TLLSRFDL---IYLMLDK-----------PDEQNDRRL--------ARHL 697
Db        657 DSIDIAQECVNFLKRQNIGVATFIGLDKMAVWAKKMTEIQTPENTPRLFDLVKVKDEKIR 716
Qy        698 VALHYENYEVSKQDALDLQTLTAYITYARQHVHPTLSDEAAEDLINGYVEMRQKGNFPGS 757
Db        717 QAFYFALRDTLVADNLDQATRVAYQKDRRWRV-VTLQGQIIE----------QSGTMTGG 765
Qy        758 SKKV-----------------ITATPRQL-----------ESMIRISEALARMRFSE--- 786
Db        766 GSKVMKGRMGSSLVIEISEEEVNKMESQLQNDSKKAMQIQEQKVQLEERVVKLRHSEREM 825
Qy        787 --VVEKVDAA-----EAVRLLDVALQQSATDHATGTIDMDLITTGVSASERIRRANLLAA 839
Db        826 RNTLEKFTASIQRLIEQEEYLNVQVKE------------------------LEANVLAT 860
Qy        840 LRELIADKISPGSSSGLKT 858
Db        861 APDKKKQKLLEENVSAFKT 879
```

RESULT 15
US-10-957-569-31
; Sequence 31, Application US/10957569
; Publication No. US20050246785A1
; GENERAL INFORMATION:
; APPLICANT: COOK, Zhihong et al.
; TITLE OF INVENTION: PROMOTER, PROMOTER CONTROL ELEMENTS, AND COMBINATIONS, AND USES
; TITLE OF INVENTION: THEREOF
; FILE REFERENCE: 2750-1577PUS3

```
;  FILE REFERENCE:
;  CURRENT APPLICATION NUMBER: US/10/467,657
;  CURRENT FILING DATE:  2003-08-11
;  PRIOR APPLICATION NUMBER: GB-0103424.8
;  PRIOR FILING DATE: 2001-02-12
;  NUMBER OF SEQ ID NOS: 9218
;  SOFTWARE: SeqWin99, version 1.04
; SEQ ID NO 84
;   LENGTH: 2004
;   TYPE: PRT
;   ORGANISM: Neisseria gonorrhoeae
US-10-467-657-84

  Query Match             2.8%;  Score 128.5;  DB 6;  Length 2004;
  Best Local Similarity   20.5%;  Pred. No. 0.48;
  Matches  200;  Conservative  125;  Mismatches  400;  Indels  249;  Gaps   47;

Qy         17 PSQSEGVSTPLPQVTSPSFDNAASPVAGRRAVRQTPTSAVRRRGRETDSARRRRSRSRS- 75
Db        443 PVAGKRPDTVLP-VLNPQVAESAGRVSPKKRMADAAADFTRRL-----AADRRRPEKAGV 496

Qy         76 -LGNSVYSSPYD--------AGTPGTPGTPVATPVYATPVGTPMGTPSFHRGTPQYKQRS 126
Db        497 PLGGGEYRFEHTDRRHIDALAGVPGRPGKGGMPEEFADMAG-----PSNSDGLVSDGRRY 551

Qy        127 ELGSQGKPLHRRRRSQS--REPGH--------RSPS------REPSADGRPSESAEP--- 167
Db        552 LKGREAETLRAGGLSEAVPSEPGRDYRPTQEARAPAKVMARPRDAAADGKPAGRAQPARA 611

Qy        168 -DDTLGGEYAYVWGTNVNIP--DVLRAIR----RFLHNYRSSAHDLNSKYIQIIEETVER 220
Db        612 KDTPVAGKAAAAKNAATEKPSSDKVRNIEAGKSRFDGGKGKSAAAQGAATEKPSEKTGKA 671

Qy        221 EEDTLNIDMSDIYDHDPDLYAKIVRYPLDIIPLLDTECQEVATSLL--PTFEKHIEARPF 278
Db        672 KPETFAKTASD----NPE-----------------EARRKARVLQGGPVYTVKERQAPQ 709

Qy        279 NLKASVHMRELNPSDIDKLV-SVKGM-----VIRCSSIIPE---------IKGAFFKCLV 323
Db        710 GFKA---LREHAESIKKRLAESIGGLAERVDVAAVSETAPDKAQMLLSQRVEGWF----- 761

Qy        324 CGHSPPLVTVVKGRVEEPTRCEK---PECAARNAMSLIHNRCTFAN------KQIVR--- 371
Db        762 -----------DGRTGKITLVAENLTPERAVWAAWHELGHRGFAADGFAKYREELERADG 810

Qy        372 ---LQETPDAIPEGE--TPHTVSMCLYNTMVDAVKPGDRIEVT----GVFKAMAVRVGPN 422
Db        811 NGLIRRIADAVQEGREGTGDAAASVRPAAVEEAVAELYAAQRTGGWAGIENRYGVKVGNG 870

Qy        423 QR--------TLRALYKTYIDCVHVKKSDRGRLQTEDPMEM--DKENDM-------YAGY 465
Db        871 LKRGIAGVLARIGALLRRVLQRLAGKAG--GAMSDADVFAMLADLHGNVEGARDAPWGGN 928

Qy        466 HES-----DTSEAANEAKIQKLKELSKLPGIYDRLSRSLAPSIWELEDIKKGLLCQLFGG 520
Db        929 HRAVMFARAEDGAAERSKSESLEKLRRAETI--RISGREVPEGGNLREYKRNAL------ 980

Qy        521 KAKKIPSGASFRGDINVLLVGDPGTSKSQLLQYVHKIAPRGIYTSGRGSSAVGLTAYVTK 580
Db        981 -----EYGKSLRG------------------PYVNKDTGREI-SLGRS----GITEILRH 1012

Qy        581 DPETRETVLESGALVLSDRGICCIDEFDKMSDNARSMLHEVMEQQTVSVAKGGIIASLNA 640
Db       1013 DYKDAEHLQSIAAIPQIIENAVYIDTLPN-EDLAKNGDIQGYEYYVSGLNVGG--ADYTV 1069

Qy        641 RTSVLACANPSGSRY----------NARLSVIDNIQLPPTLLSRFDLIYLMLDKPDEQND 690
Db       1070 RAAVAVSRN--GNRYYDHKLTKIEKGNLLSLLDRVSTTGASESKSPLSGI--------DD 1119

Qy        691 RRLARHLV--------ALHYENYEVSKQDALDLQTLTAYITYARQHVHPTLSD------- 735
Db       1120 KRLLQILQDKDAGKGGIADFDTEAVRFSRAANIEAAIGRITGKKSDLRNALKDRWDASKG 1179

Qy        736 --------EAAEDLINGYVE-MRQKGN----FPGSSKKVITATPRQLESMIRISE----A 778
Db       1180 IQLQFLGRRQIEDIYGGVLDGLKEYGRLSELFGADANKAVTEADKVVREWGRLKEEDAKA 1239

Qy        779 LARMRFSEVVEKVDA-----AEAVRLLDVALQQSATDHATGTIDMDLITTGVSASERIRR 833
Db       1240 LADLMHDATLAKVDADPLMRKDAQKRLDGI--RTALDIADGKIEKAEAAV-ASAGARIAR 1296

Qy        834 ANLLAALRELIADK 847
Db       1297 ADAAYNKAQRAADK 1310

RESULT 13
US-10-467-657-6322
; Sequence 6322, Application US/10467657
; Publication No. US20050260581A1
; GENERAL INFORMATION:
;  APPLICANT: CHIRON SpA
;  APPLICANT: FONTANA Maria Rita
;  APPLICANT: PIZZA Mariagrazia
;  APPLICANT: MASIGNANI Vega
;  APPLICANT: MONACI Elisabetta
;  TITLE OF INVENTION: GONOCOCCAL PROTEINS AND NUCLEIC ACIDS
;  FILE REFERENCE:
;  CURRENT APPLICATION NUMBER: US/10/467,657
;  CURRENT FILING DATE:  2003-08-11
;  PRIOR APPLICATION NUMBER: GB-0103424.8
;  PRIOR FILING DATE: 2001-02-12
;  NUMBER OF SEQ ID NOS: 9218
;  SOFTWARE: SeqWin99, version 1.04
; SEQ ID NO 6322
;   LENGTH: 2004
;   TYPE: PRT
;   ORGANISM: Neisseria gonorrhoeae
US-10-467-657-6322

  Query Match             2.8%;  Score 128.5;  DB 6;  Length 2004;
  Best Local Similarity   20.5%;  Pred. No. 0.48;
  Matches  200;  Conservative  125;  Mismatches  400;  Indels  249;  Gaps   47;

Qy         17 PSQSEGVSTPLPQVTSPSFDNAASPVAGRRAVRQTPTSAVRRRGRETDSARRRRSRSRS- 75
Db        443 PVAGKRPDTVLP-VLNPQVAESAGRVSPKKRMADAAADFTRRL-----AADRRRPEKAGV 496

Qy         76 -LGNSVYSSPYD--------AGTPGTPGTPVATPVYATPVGTPMGTPSFHRGTPQYKQRS 126
Db        497 PLGGGEYRFEHTDRRHIDALAGVPGRPGKGGMPEEFADMAG-----PSNSDGLVSDGRRY 551

Qy        127 ELGSQGKPLHRRRRSQS--REPGH--------RSPS------REPSADGRPSESAEP--- 167
Db        552 LKGREAETLRAGGLSEAVPSEPGRDYRPTQEARAPAKVMARPRDAAADGKPAGRAQPARA 611

Qy        168 -DDTLGGEYAYVWGTNVNIP--DVLRAIR----RFLHNYRSSAHDLNSKYIQIIEETVER 220
Db        612 KDTPVAGKAAAAKNAATEKPSSDKVRNIEAGKSRFDGGKGKSAAAQGAATEKPSEKTGKA 671

Qy        221 EEDTLNIDMSDIYDHDPDLYAKIVRYPLDIIPLLDTECQEVATSLL--PTFEKHIEARPF 278
Db        672 KPETFAKTASD----NPE-----------------EARRKARVLQGGPVYTVKERQAPQ 709

Qy        279 NLKASVHMRELNPSDIDKLV-SVKGM-----VIRCSSIIPE---------IKGAFFKCLV 323
Db        710 GFKA---LREHAESIKKRLAESIGGLAERVDVAAVSETAPDKAQMLLSQRVEGWF----- 761

Qy        324 CGHSPPLVTVVKGRVEEPTRCEK---PECAARNAMSLIHNRCTFAN------KQIVR--- 371
Db        762 -----------DGRTGKITLVAENLTPERAVWAAWHELGHRGFAADGFAKYREELERADG 810

Qy        372 ---LQETPDAIPEGE--TPHTVSMCLYNTMVDAVKPGDRIEVT----GVFKAMAVRVGPN 422
Db        811 NGLIRRIADAVQEGREGTGDAAASVRPAAVEEAVAELYAAQRTGGWAGIENRYGVKVGNG 870
```

```
Db        705 AGELAVGREILKAGSIGVGGVDYVLSCILGNQAESNNWQELLGRLCLI----------D 753

Qy        691 RRLAR-------HLVALHYENYEVSKQDALDL--QTLTAYITYARQHVHPTLSDEAAEDL 741

Db        754 RLLLEISAEFYPHIVST-----DVSQAEPVEIRYKKLLSLLAFALQSI---------DN 798

Qy        742 INGYVEMRQKGNFPGSSKKVITATP--RQLESMIRISE----ALARMRFSEVVEKVDAAE 795

Db        799 SHSMVGKLSRRIYLSSARMVTTVPPLFSKLVTMLSASGSSHFARMRRRLMAIADEVEIAE 858

Qy        796 AVRL-----LD------VAL------QQSATDHATGTIDMDLITTGVSASERIRRANLLA 838

Db        859 VIQLGSEDTLDGQQDSSQALAPPRYPESSSLEH---------TAHVEKTGKGLKATRLS 908

Qy        839 ALRELIADKISPGSSSGLKTSQLLE 863

Db        909 ASSEDISDRLA-GVSVGLPSSATTE 932
```

RESULT 10
US-11-074-176-48
; Sequence 48, Application US/11074176
; Publication No. US20050250135A1
; GENERAL INFORMATION:
; APPLICANT: Klaenhammer, Todd R.
; APPLICANT: Russell, William M.
; APPLICANT: Altermann, Eric
; APPLICANT: McAuliffe, Olivia
; APPLICANT: Peril, Andrea Azcarate
; TITLE OF INVENTION: Nucleic Acid Sequences Encoding
; TITLE OF INVENTION: Stress-Related Proteins and Uses Therefore
; FILE REFERENCE: 5051-694
; CURRENT APPLICATION NUMBER: US/11/074,176
; CURRENT FILING DATE: 2005-03-07
; PRIOR APPLICATION NUMBER: 60/551,161
; PRIOR FILING DATE: 2004-03-08
; NUMBER OF SEQ ID NOS: 381
; SOFTWARE: FastSEQ for Windows Version 4.0
; SEQ ID NO 48
; LENGTH: 420
; TYPE: PRT
; ORGANISM: Lactobacillus acidophilus
US-11-074-176-48

```
  Query Match             2.8%;  Score 129;  DB 7;  Length 420;
  Best Local Similarity   22.2%;  Pred. No. 0.042;
  Matches  103;  Conservative   77;  Mismatches  164;  Indels  120;  Gaps   23;

Qy        433 YIDCVHVKKSDRGRLQTEDPMEMDKENDMYAG---YHESDTSEAA-----NEAKIQKLKE 484

Db          5 FTDQEEIKCAFCGKTQ-------DQVKKMIAGNGVYICNECVDLAKKIIDDELRADSLKT 57

Qy        485 LSKLPGIYDRLSRSLAPSIWELEDIKKGLLCQLFG-GKAKKIPSGASF-------RGDI- 535

Db         58 ASELP---------------KPVEIKKQLDQYVIGQDRAKKVLSVAVYNHYKRISQMDVD 102

Qy        536 --------NVLLVGDPGTSKSQLLQYVHKI--APRGIYTSGRGSSAVGLT--AYVTKDPE 583

Db        103 SSTELQKSNIAMIGPTGSGKTYLAQTLARILNVPFAI------ADATTLTEAGYVGEDVE 156

Qy        584 TRETVLESGA---LVLSDRGICCIDEFDKMSDNA------RSMLHEVMEQQTVSVAKGGI 634

Db        157 NILLKLLQNADYDLERAQRGIIYIDEIDKISKKSENVSITRDVSGEGVQQSLLKILEG-- 214

Qy        635 IASLNARTSVLACANPSGSRYNARLSVIDNIQLPPTLLSRFDLIYLM---LDKPDEQNDR 691

Db        215 ---------TIASVPPQGGRKHPQQEMIK--------MDTTNILFIVGGAFDGIEQIVKS 257

Qy        692 RLARHLVALHYENYEVSKQDALDLQTLTAYITYA-------------RQHVHPTLSDEAA 738

Db        258 RLGKKTIGFGAEN-EVNKVDADD---WTRHLTTADLVKFGMIPEFIGRIPIITTLDKLDN 313

Qy        739 EDLINGYVEMRQKGNFPGSSKKVITATPRQLESMIRISEALARMRFSEVVEKVDAAEAVR 798

Db        314 KDLVRVLTE--PKNALVKQYKKLLSLDGVELKFTDGALKAIADL----AIQRNMGARGLR 367

Qy        799 ------LLDVALQ-QSATDHATGTIDMDLITTGVSASERIRRAN 835

Db        368 TIIENSIMDIMYETPSEEDIESVEVTKDVITR--HAQPRITRKN 409
```

RESULT 11
US-10-467-657-7810
; Sequence 7810, Application US/10467657
; Publication No. US20050260581A1
; GENERAL INFORMATION:
; APPLICANT: CHIRON SpA
; APPLICANT: FONTANA Maria Rita
; APPLICANT: PIZZA Mariagrazia
; APPLICANT: MASIGNANI Vega
; APPLICANT: MONACI Elisabetta
; TITLE OF INVENTION: GONOCOCCAL PROTEINS AND NUCLEIC ACIDS
; FILE REFERENCE:
; CURRENT APPLICATION NUMBER: US/10/467,657
; CURRENT FILING DATE: 2003-08-11
; PRIOR APPLICATION NUMBER: GB-0103424.8
; PRIOR FILING DATE: 2001-02-12
; NUMBER OF SEQ ID NOS: 9218
; SOFTWARE: SeqWin99, version 1.04
; SEQ ID NO 7810
; LENGTH: 498
; TYPE: PRT
; ORGANISM: Neisseria gonorrhoeae
US-10-467-657-7810

```
  Query Match             2.8%;  Score 129;  DB 6;  Length 498;
  Best Local Similarity   20.4%;  Pred. No. 0.055;
  Matches   70;  Conservative   49;  Mismatches  122;  Indels  102;  Gaps   11;

Qy        514 LCQLFGGKAKKIPSGASFRGDINVLLVGDPGTSKSQLLQYVHKIAP------------- 559

Db        193 LCDVKGQHSARLALEIAAAGGHSLLMMGPPGTGKSMLSQRLPGILPPLAEDELVEVWALR 252

Qy        560 ------------RGIYTSGRGSSAVGLTAYVTKDPETRETVLESGALVLSDRGICCIDE 606

Db        253 SLLPNHQQQLDSNRPFRSPHHGASAAAMVG-GGSDP-------RPGEISLAHHGVLFLDE 304

Qy        607 FDKMSDNARSMLHEVMEQQTVSVAKGGIIASLNARTSVLACANP--------------- 650

Db        305 LPEFDRKVLEVLREPLENGEIHISRAARQAVYPAKFQLVAAMNPCPCGYLGHPVKPCRCT 364

Qy        651 --SGSRYNARLSVIDNIQLPPTLLSRFDLIYLMLDKPDEQNDRRLARHLVALHYENYEVS 708

Db        365 PESVARYRSKIS--------GPLLDRID---LTIEVPS----------LSAAELMQQEAG 403

Qy        709 KQDALDLQTLTAY--ITYARQ-HVHPTLS-----------DEAAEDLINGYVEMRQKGNF 754

Db        404 ESSASVLERVIAARGKQYARQGKVNAALSVSELDSQACIQKEAQEAL------------ 450

Qy        755 PGSSKKVITATPRQLESMIRISEALARMRFSEVVEKVDAAEAV 797

Db        451 -GSLLEKLSLSARSFHRIMRVARTLADLAGDEEVGRSHVMKAI 492
```

RESULT 12
US-10-467-657-84
; Sequence 84, Application US/10467657
; Publication No. US20050260581A1
; GENERAL INFORMATION:
; APPLICANT: CHIRON SpA
; APPLICANT: FONTANA Maria Rita
; APPLICANT: PIZZA Mariagrazia
; APPLICANT: MASIGNANI Vega
; APPLICANT: MONACI Elisabetta
; TITLE OF INVENTION: GONOCOCCAL PROTEINS AND NUCLEIC ACIDS

```
Db        168 ---------------------LHKME------------------DRPEE---------- 177

Qy        234 DHDPDLYAKIVRYPLDIIPLLDTECQEVATSLLPTFEKH--IEARPFNLKASVHMRELNP 291
Db        178 --------RMIR-------------EKLKATCMPAW-KHEWLERR--NRRGPV------- 206

Qy        292 SDIDKLVSVKGMVIRCSSIIPEIKGAFFKCLVCGHSPPLVTVVKGRVEEPT-------RC 344
Db        207 --VVKPIPIKGDGSEVNNLAAEPQGE-------GQAGSAAPAPKGR-RSPSPGSSPSGRS 256

Qy        345 EKPEC--AARNAMSLIHNRCTFANKQIVRLQETPDAIPEGETPHTVSMCLYNTMVDAVKP 402
Db        257 VKPESPGVRRKRVSPV----PFQSGRITPPRRAPS--PDGFSPYSP-------------- 296

Qy        403 GDRIEVTG--VFKAMAVR------VGPNQRTLRA-----LYKTYID-----------CVH 438
Db        297 ----EETSRRVNKVMRARLYLLQQIGPNSFLIGGDSPDNKYRVFIGPQNCSCGRGAFCIH 352

Qy        439 V--KKSDRGRLQTEDPMEMDKE------NDMYAGYH------------------------ 466
Db        353 LLFVMLRVFQLEPSDPMLWRKTLKNFEVESLFQKYHSRRSSRIKAPSRNTIQKFVSRMSN 412

Qy        467 -----ESDTSEAANEAKIQKLKELS---KLPGIYDRLSRSLAP------------SIW-- 504
Db        413 SHTLSSSSTSTSSSENSIKDEEEQMCPICLLGMLDEESLTVCEDGCRNKLHHHCMSIWAE 472

Qy        505 ELEDIKKGLLCQLFGGK-----------AKKIPSGASFR--------------------- 532
Db        473 ECRRNREPLICPLCRSKWRSHDFYSHELSSPVESPASLRAVQQPSSPQQPVAGSQRRNQE 532

Qy        533 GDINVLLVGD---PGTSKS--------------------------QLLQYVHKIA---- 558
Db        533 SSFNLTHFGTQQIPSAYKDLAEPWIQVFGMELVGCLFSRNWNVREMALRRLSHDVSGALL 592

Qy        559 -----PRGIYTSGRGSSAVGLTAYVTKDPETRETVLESGALVLSDRGICCIDEFDKMSDN 613
Db        593 LANGESTGNSGGGSGGSLSAGAASGSSQPSISGDVVEACCSVLS---IVCADPVYKVYVA 649

Qy        614 A----RSML-----HEVMEQ-----------QTVSVAKGGIIASLNARTSVLA------- 646
Db        650 ALKTLRAMLVYTPCHSLAERIKLQRLLRPVVDTILVK----CADANSRTSQLSISTVLEL 705

Qy        647 CANPSGSRYNAR----------------LSVIDNIQLPPT----LLSRFDLI-YLMLDKP 685
Db        706 CKGQAGELAVGREILKAGSIGVGGVDYVLSCILGNQAESNNWQELLGRLCLIDRLLLEFP 765

Qy        686 DEQNDRRLARHLVALHYENYEVSKQDALDL--QTLTAYIYTYARQHVHPTLSDEAAEDLIN 743
Db        766 AE-----FYPHIVST-----DVSQAEPVEIRYKKLLSLLTFALQSIDNSHS-------MV 808

Qy        744 GYVEMRQKGNFPGSSKKVITATPRQLESMIRISEALA-------RMRFSEVVEKVDAAEA 796
Db        809 GKLSRR----IYLSSARMVTAVPAVFSKLVTMLNASGSTHFTRMRRRLMAIADEVEIAEV 864

Qy        797 VRL-----LD------------VALQQSATDHATGTIDMDLITTGVSASERIRRANLLAA 839
Db        865 IQLGVEDTVDGHQDSLQAVAPTSCLENSSLEH---TVHREKTGKGLSATR-------LSA 914

Qy        840 LRELIADKISPGSSSGLKTSQLLE 863
Db        915 SSEDISDRLA-GVSVGLPSSTTTE 937
```

RESULT 9
US-11-004-057-21
; Sequence 21, Application US/11004057
; Publication No. US20050244846A1
; GENERAL INFORMATION:
; APPLICANT: Johnson, Gary L.
; TITLE OF INVENTION: MEKK1 PROTEINS AND FRAGMENTS THEREOF FOR USE IN REGULATING
; TITLE OF INVENTION: APOPTOSIS
; FILE REFERENCE: CPI-042CPPC
; CURRENT APPLICATION NUMBER: US/11/004,057
; CURRENT FILING DATE: 2004-12-02
; PRIOR APPLICATION NUMBER: US/09/403,075
; PRIOR FILING DATE: 2000-05-10
; NUMBER OF SEQ ID NOS: 21
; SOFTWARE: PatentIn Ver. 2.0
; SEQ ID NO 21
;   LENGTH: 1493
;   TYPE: PRT
;   ORGANISM: Rattus norvegicus
US-11-004-057-21

```
  Query Match             3.2%;  Score 145.5;  DB 7;  Length 1493;
  Best Local Similarity   19.7%;  Pred. No. 0.02;
  Matches  218;  Conservative  120;  Mismatches  334;  Indels  433;  Gaps   57;

Qy         11 AVSSPYPSQSEGVSTPLPQV--TSPSFDNAASPVAGRRAVRQTPTSAVRRRGRETDSARR 68
Db          9 ASSSGFP----GAAAASPEAGGGGGALQGSGAPAAGAGLLRETGSA-----GRERADWRR 59

Qy         69 ---RRSRSRSLGNSVYSSPYDAGTPGTPGTPVATPVYATPVGTPMGTPSFH--RGTPQYK 123
Db         60 QQLRKVRSVELDQLPEQPLFLTASPPCPSTSPS----PEPADAAAGASGFQPAAGPPPPG 115

Qy        124 QRSELGSQGKPLHRRRRSQSREPGHRSPSREPSADGRPSESAEPDDTLGGEYAYVWGTNV 183
Db        116 AASRCGSHSAELAAARDSGARSPAGAEPPSAAAPSGREMENKE---TLKG---------- 162

Qy        184 NIPDVLRAIRRFLHNYRSSAHDLNSKYIQIIEETVEREEDTLNIDMSDIYDHDPDLYAKI 243
Db        163 ------------------LHKMDDR----PEERMIREK--------------------- 178

Qy        244 VRYPLDIIPLLDTECQEVATSLLPTFEKH--IEARPFNLKASVHMRELNPSDIDKLVSVK 301
Db        179 ----------LKATC-------MPAW-KHEWLERR--NRRGPV---------VVKPIPIK 209

Qy        302 GMVIRCSSIIPEIKGAFFKCLVCGHSPPLVTVVKGRVEEPT-------RCEKPEC--AAR 352
Db        210 GDGSEMSNLAAELQGE-------GQAGSAAPAPKGR-RSPSPGSSPSGRSGKPESPGVRR 261

Qy        353 NAMSLIHNRCTFANKQIVRLQETPDAIPEGETPHTVSMCLYNTMVDAVKPGDRIEVTG-- 410
Db        262 KRVSPV----PFQSGRITPPRRAPS--PDGFSPYSP------------------EETSRR 297

Qy        411 VFKAMAVR------VGPNQRTLRA-----LYKTYID-----------CVHV--KKSDRGR 446
Db        298 VNKVMRARLYLLQQIGPNSFLIGGDSPDNKYRVFIGPQNCSCGRGTFCIHLLFVMLRVFQ 357

Qy        447 LQTEDPMEMDKE------NDMYAGYH---------------------------ESDTS 471
Db        358 LEPSDPMLWRKTLKNFEVESLFQKYHSRRSSRIKAPSRNTIQKFVSRMSNCHTLSSSSTS 417

Qy        472 EAANEAKIQKLKELS---KLPGIYDRLSRSLAP------------SIW--ELEDIKKGLL 514
Db        418 TSSSENSIKDEEEQMCPICLLGMLDEESLTVCEDGCRNKLHHHCMSIWAEECRRNREPLI 477

Qy        515 CQLFGGK-----------AKKIPSGASFRGDINVLLVGDPGT--------------SKSQ 549
Db        478 CPLCRSKWRSHDFYSHELSSPVDSPTSLRG------VQQPSSPQQPVAGSQRRNQESNFN 531

Qy        550 LLQY----------------------------------VHKIAPR--------------- 560
Db        532 LTHYGTQQIPPAYKDLAEPWIQAFGMELVGCLFSRNWNVREMALRRLSHDVSGALLLANG 591

Qy        561 ---GIYTSGRGSSAVGLTAYVTKDPETRETVLESGALVLSDRGICCIDEFDKMSDNA--- 614
Db        592 ESTGTSGGGSGGSLSAGAASGSSQPSISGDVVEAFCSVLS---IVCADPVYKVYVAALKT 648

Qy        615 -RSML-----HEVMEQ-----------QTVSVAKGGIIASLNARTSVLA-------CANP 650
Db        649 LRAMLVYTPCHSLAERIKLQRLLRPVVDTILVK----CADANSRTSQLSISTLLELCKGQ 704

Qy        651 SGSRYNAR----------------LSVIDNIQLPPT----LLSRFDLIYLMLDKPDEQND 690
```

```
Qy        805 -------------QQSATD------------------HATGTIDMDLITTGVSASERIR 832
Db        683 EDQEQKRKRRKTRQPDAKDGDSYDPYDFSDTEEEMPQVHTPKTAD----SQETKESQKVE 738

Qy        833 RA-NLLAALRELIADKISPGSSSGLKTSQLLEDIRSQSSVDVSLQDIKNALGSLQGEGFL 891
Db        739 LSESRLKAFKVALLDVFREAHAQSIGMNRLTESINRDSEEPFSSVEIQAALSKMQDDNQV 798

Qy        892 TVHGDIV 898
Db        799 MVSEGII 805
```

RESULT 7
US-10-507-275-3
; Sequence 3, Application US/10507275
; Publication No. US20050250166A1
; GENERAL INFORMATION:
; APPLICANT: Masai, Hisao
; APPLICANT: Tamai, Katsuyuki
; APPLICANT: Medical and Biological Laboratories Co., Ltd.
; APPLICANT: Japan Science and Technology Agency
; APPLICANT: Ginkgo Biomedical Research Institute Co., Ltd.
; TITLE OF INVENTION: Cdc7-ASK Kinase Complex, Substrates of the Kinase Complex,
; TITLE OF INVENTION: Specific Antibodies to the Substrates, and Screening Methods
; TITLE OF INVENTION: Using the Same to Screen for Compounds Comprising Cdc7-ASK
; TITLE OF INVENTION: Kinase Inhititory Ability
; FILE REFERENCE: 082368-001100US
; CURRENT APPLICATION NUMBER: US/10/507,275
; CURRENT FILING DATE: 2004-09-09
; PRIOR APPLICATION NUMBER: JP 2002-067702
; PRIOR FILING DATE: 2002-03-12
; PRIOR APPLICATION NUMBER: WO PCT/JP03/02918
; PRIOR FILING DATE: 2003-03-12
; NUMBER OF SEQ ID NOS: 21
; SOFTWARE: PatentIn Ver. 2.1
; SEQ ID NO 3
; LENGTH: 892
; TYPE: PRT
; ORGANISM: Homo sapiens
US-10-507-275-3

Query Match 18.1%; Score 835; DB 6; Length 892;
Best Local Similarity 27.5%; Pred. No. 4.7e-51;
Matches 247; Conservative 151; Mismatches 348; Indels 152; Gaps 24;

```
Qy         38 AASPVAGRRAVRQTPTSAVRRRGRETDSARRRRSRSRSLGNSVYSSPYDAGTPGTPGTPV 97
Db          2 ASSP-AQRRRGNDPLTSSPGRSSRRTDAL--TSSPGRDL------PPFEDESEGLLGT-- 50

Qy         98 ATPVYATPVGTPM---GTPSFHRGTPQY------------------KQRSE-------- 127
Db         51 EGPLEEEEDGEELIGDGMERDYRAIPELDAYEAEGLALDDEDVEELTASRREAADGPCGT 110

Qy        128 -------LG--SQGKPLHRRRRSQSREPGHRSP---SREPSADGRPSESAEPDDTLGGEY 175
Db        111 VTGSWPGLGACAVGSCMTAMRRTRSALPASAASGAGTEDGEEDEQMIESIENLEDLKGHS 170

Qy        176 AYVW----GTNVNIPDVLRAIRRFLHNY-RSSAHDLNSKYIQIIEETVEREEDTLNIDMS 230
Db        171 VREWVSMAGPRL---EIHHRFKNFLRTHVDSHGHNV---FKERISDMCKENRESLVVNYE 224

Qy        231 DIYDHDPDLYAKIVRYPLDIIPLLDTECQEVATSLLPTFEK---HIEARPFNLKASVHMR 287
Db        225 DLAAREHVLAYFLPEAPAELLQIFDEAALEVVLAMYPKYDRITNHIHVRISHLPLVEELR 284

Qy        288 ELNPSDIDKLVSVKGMVIRCSSIIPEIKGAFFKCLVCGHSPPLVTVVKGRVEEPTRCEKP 347
Db        285 SLRQLHLNQLIRTSGVVTSCTGVLPQLSMVKYNCNKCNFVLGPFCQSQNQEVKPGSC--P 342

Qy        348 ECAARNAMSLIHNRCTFANKQIVRLQETPDAIPEGETPHTVSMCLYNTMVDAVKPGDRIE 407
Db        343 ECQSAGPFEVNMEETIYQNYQRIRIQESPGKVAARRLPRSKDAILLADLVDSCNAGDEIE 402

Qy        408 VTGVFKAMAVRVGPNQRTLRA--LYKTYIDCVHVKKSDR----GRLQTEDPMEMDKENDM 461
Db        403 LTGIYHN---NYDGSLNTANGFPVFATVILANHVAKKDNKVAVGELTDED---------- 449

Qy        462 YAGYHESDTSEAANEAKIQKLKELSKLPGIYDRLSRSLAPSIWELEDIKKGLLCQLFGGK 521
Db        450 -----------------VKMITSLSKDQQIGEKIFASIAPSIYGHEDIKRGPALALFGGE 492

Qy        522 AKKIPSGASFRGDINVLLVGDPGTSKSQLLQYVHKIAPRGIYTSGRGSSAVGLTAYVTKD 581
Db        493 PKNPGGKHKVRGDINVLLCGDPGTAKSQFLKYIEKVSSRAIFTTGQGASAVAVTAYVQRH 552

Qy        582 PETRETVLESGALVLSDRGICCIDEFDKMSDNARSMLHEVMEQQTVSVAKGGIIASLNAR 641
Db        553 PVSREWTLEAGALVLADRGVCLIDEFDKMNDQDRTSIHEAMEQQSISISKAGIVTSLQAR 612

Qy        642 TSVLACANPSGSRYNARLSVIDNIQLPPTLLSRFDLIYLMLDKPDEQNDRRLARHLVALH 701
Db        613 CTVIAAANPIGGRYDPSLTFSENVDLTEPIISRFDILCVVRDTVDPVQDEMLARFVVGSH 672

Qy        702 YENYEVSKQ--------------------DALDLQTLTAYITYARQHVHPTLSDEAAEDL 741
Db        673 VRHHPSNKEEEGLANGSAAEPAMPNTYGVEPLPQEVLKKYIIYAKERVHPKLNQMDQDKV 732

Qy        742 INGYVEMRQKGNFPGSSKKVITATPRQLESM-------IRISEALARMRFSEVVEKVDAA 794
Db        733 AKMYSDLRKESMATGS----IPITVRHIESMSHGGGPRAHPSAGLCDRRRRQHGHPRDAG 788

Qy        795 EAVRLLDVALQQSATDHATGTIDMDLITTGVSASERIRRAN---LLAALRELIADKIS 849
Db        789 ELHRHTEVQRHRSMRKTFARYLS-------------FRRDNNELLLFILKQLVAEQVT 833
```

RESULT 8
US-11-004-057-4
; Sequence 4, Application US/11004057
; Publication No. US20050244846A1
; GENERAL INFORMATION:
; APPLICANT: Johnson, Gary L.
; TITLE OF INVENTION: MEKK1 PROTEINS AND FRAGMENTS THEREOF FOR USE IN REGULATING
; TITLE OF INVENTION: APOPTOSIS
; FILE REFERENCE: CPI-042CPPC
; CURRENT APPLICATION NUMBER: US/11/004,057
; CURRENT FILING DATE: 2004-12-02
; PRIOR APPLICATION NUMBER: US/09/403,075
; PRIOR FILING DATE: 2000-05-10
; NUMBER OF SEQ ID NOS: 21
; SOFTWARE: PatentIn Ver. 2.0
; SEQ ID NO 4
; LENGTH: 1493
; TYPE: PRT
; ORGANISM: Mus musculus
US-11-004-057-4

Query Match 3.3%; Score 150; DB 7; Length 1493;
Best Local Similarity 19.6%; Pred. No. 0.0096;
Matches 216; Conservative 126; Mismatches 336; Indels 426; Gaps 58;

```
Qy         11 AVSSPYPSQSEGVSTPLPQV-----TSPSFDNAASPVAGRRAVRQTPTSAVRRRGRETDS 65
Db          9 ASSSGFP----GAAAASPEAGGGGGGGALQGSGAPAAGAAGLLREPGSA----GRERAD 60

Qy         66 ARRRRSRSRSLGNSVYSSPYD----------AGTPGTPGTPVATPVYATPVGTPMGTPSF 115
Db         61 WRRRHVR------KVRSVELDQLPEQPLFLAAASPPCPSTSPS----PEPADAAAGASRF 110

Qy        116 H--RGTPQYKQRSELGSQGKPLHRRRRSQSREPGHRSPSREPSADGRPSESAEPDDTLGG 173
Db        111 QPAAGPPPPGAASRCGSHSAELAAARDSGARSPAGAEPPSAAAPSGREMENKE---TLKG 167

Qy        174 EYAYVWGTNVNIPDVLRAIRRFLHNYRSSAHDLNSKYIQIIEETVEREEDTLNIDMSDIY 233
```

```
Qy        882 L 882

Db        773 I 773

RESULT 5
US-11-087-227-16
; Sequence 16, Application US/11087227
; Publication No. US20050260566A1
; GENERAL INFORMATION:
;  APPLICANT: Fischer, Timothy J.
;  APPLICANT:  Malinowski, Douglas P.
;  APPLICANT:  Taylor, Adriann J.
;  APPLICANT:  Parker, Margaret R.
;  TITLE OF INVENTION: METHODS AND COMPOSITIONS FOR THE
;  TITLE OF INVENTION:  DETECTION OF CERVICAL DISEASE
;  FILE REFERENCE: 046143/287139
;  CURRENT APPLICATION NUMBER: US/11/087,227
;  CURRENT FILING DATE:  2005-03-23
;  PRIOR APPLICATION NUMBER: 60/556,495
;  PRIOR FILING DATE: 2004-03-24
;  NUMBER OF SEQ ID NOS: 90
;  SOFTWARE: FastSEQ for Windows Version 4.0
; SEQ ID NO 16
;   LENGTH: 543
;   TYPE: PRT
;   ORGANISM: Homo sapiens
US-11-087-227-16

  Query Match             18.9%;  Score 870;  DB 7;  Length 543;
  Best Local Similarity   37.0%;  Pred. No. 8e-54;
  Matches  200;  Conservative  103;  Mismatches  176;  Indels   62;  Gaps   13;

Qy        317 AFFKCLVCG---HSPPLVTVVKGRVEEPT-----RCEKPECAARNA---MSLIHNRCTFA 365

Db          4 ATYTCDQCGAETYQP---------IQSPTFMPLIMCPSQECQTNRSGGRLYLQTRGSRFI 54

Qy        366 NKQIVRLQETPDAIPEGETPHTVSMCLYNTMVDAVKPGDRIEVTGVFKAMAVRVGPNQRT 425

Db         55 KFQEMKMQEHSDQVPVGNIPRSITVLVEGENTRIAQPGDHVSVTGIFLPI-LRTGFRQVV 113

Qy        426 LRALYKTYIDCVHVKKSDRGRLQTEDPMEMDKENDMYAGYHESDTSEAANEAKIQKLKEL 485

Db        114 QGLLSETYLEAHRIVK-------------MNKSEDDESG---------AGELTREELRQI 151

Qy        486 SKLPGIYDRLSRSLAPSIWELEDIKKGLLCQLFGGKAKKIPSGASFRGDINVLLVGDPGT 545

Db        152 AE-EDFYEKLAASIAPEIYGHEDVKKALLLLLVGG-VDQSPRGMKIRGNINICLMGDPGV 209

Qy        546 SKSQLLQYVHKIAPRGIYTSGRGSSAVGLTAYVTKDPETRETVLESGALVLSDRGICCID 605

Db        210 AKSQLLSYIDRLAPRSQYTTGRGSSGVGLTAAVLRDSVSGELTLEGGALVLADQGVCCID 269

Qy        606 EFDKMSDNARSMLHEVMEQQTVSVAKGGIIASLNARTSVLACANPSGSRYNARLSVIDNI 665

Db        270 EFDKMAEADRTAIHEVMEQQTISIAKAGILTTLNARCSILAAANPAYGRYNPRRSLEQNI 329

Qy        666 QLPPTLLSRFDLIYLMLDKPDEQNDRRLARHLVALHYENYEVSKQ-DALDLQTLTAYITY 724

Db        330 QLPAALLSRFDLLWLIQDRPDRDNDLRLAQHITYVHQHSRQPPSQFEPLDMKLMRRYIAM 389

Qy        725 ARQHVHPTLSDEAAEDLINGYVEMRQKGNFPGSSKKVITATPRQLESMIRISEALARMRF 784

Db        390 CREK-QPMVPESLADYITAAYVEMRREA---WASKDATYTSARTLLAILRLSTALARLRM 445

Qy        785 SEVVEKVDAAEAVRLLDVALQQSATDHATGTIDMDLITTGVSASERIRRANLLAALRELI 844

Db        446 VDVVEKEDVNEAIRLMEMSKDSLLGD------------KGQTARTQRPADVIFATVRELV 493

Qy        845 A 845

Db        494 S 494

RESULT 6
US-10-878-556A-80
; Sequence 80, Application US/10878556A
; Publication No. US20050266399A1
; GENERAL INFORMATION:
;  APPLICANT: Hoffmann La-Roche Inc.
;  TITLE OF INVENTION: HCV regulated protein expression
;  FILE REFERENCE: 21762
;  CURRENT APPLICATION NUMBER: US/10/878,556A
;  CURRENT FILING DATE:  2004-06-28
;  NUMBER OF SEQ ID NOS: 199
;  SOFTWARE: PatentIn version 3.1
; SEQ ID NO 80
;   LENGTH: 808
;   TYPE: PRT
;   ORGANISM: Homo sapiens
;  PUBLICATION INFORMATION:
;   DATABASE ACCESSION NUMBER: sw_hum/mcm3_human
;   DATABASE ENTRY DATE: 1992-05-01
US-10-878-556A-80

  Query Match             18.3%;  Score 845;  DB 6;  Length 808;
  Best Local Similarity   27.4%;  Pred. No. 8.1e-52;
  Matches  232;  Conservative  149;  Mismatches  294;  Indels  172;  Gaps   18;

Qy        180 GTNVNIPDV-LRAIRRFLHNYRSSAHDLNSKYIQIIEETVEREEDTLNIDMSDIYDHDPD 238

Db          3 GT-VVLDDVELREAQRDYLDFLDDEED-QGIYQSKVRELISDNQYRLIVNVNDLRRKNEK 60

Qy        239 LYAKIVRYPLDIIPLLDTECQEVATSLLPTFEKHIEARPFNLKASVHMRELNPSDIDK-- 296

Db         61 RANRLLNNAFEELVAFQRALKDFVASIDATYAKQYEEFYVGLEGSFGSKHVSPRTLTSCF 120

Qy        297 ---LVSVKGMVIRCSSIIPEIKGAFFKCLVCGHSPPLVTVVKGRVEE------------- 340

Db        121 LSCVVCVEGIVTKCSLVRPKVVRSVHYC------PATKKTIERRYSDLTTLVAFPSSSVY 174

Qy        341 PTRCEKPECAARNAMSLIHNRCTFANKQIVRLQETPDAIPEGETPHTVSMCLYNTMVDAV 400

Db        175 PTKDEE-----NNPLETEYGLSVYKDHQTITIQEMPEKAPAGQLPRSVDVILDDDLVDKA 229

Qy        401 KPGDRIEVTGVFKAMAVRVGPNQRTLRALYKTYIDCVHVKKSDRGRLQTEDPMEMDKEND 460

Db        230 KPGDRVQVVGTYRCLP---GKKGGYTSGTFRTVLIACNVK-------------QMSK--- 270

Qy        461 MYAGYHESDTSEAANEAKIQKLKELSKL--PGIYDRLSRSLAPSIWELEDIKKGLLCQLF 518

Db        271 --------DAQPSFSAEDIAKIKKFSKTRSKDIFDQLAKSLAPSIHGHDYVKKAILCLLL 322

Qy        519 GGKAKKIPSGASFRGDINVLLVGDPGTSKSQLLQYVHKIAPRGIYTSGRGSSAVGLTAYV 578

Db        323 GGVERDLENGSHIRGDINILLIGDPSVAKSQLLRYVLCTAPRAIPTTGRGSSGVGLTAAV 382

Qy        579 TKDPETRETVLESGALVLSDRGICCIDEFDKMSDNARSMLHEVMEQQTVSVAKGGIIASL 638

Db        383 TTDQETGERRLEAGAMVLADRGVVCIDEFDKMSDMDRTAIHEVMEQGRVTIAKAGIHARL 442

Qy        639 NARTSVLACANPSGSRYNARLSVIDNIQLPPTLLSRFDLIYLMLDKPDEQNDRRLARHLV 698

Db        443 NARCSVLAAANPVYGRYDQYKTPMENIGLQDSLLSRFDLLFIMLDQMDPEQDREISDHVL 502

Qy        699 ALH----------------------YENYEVSKQDALDLQTLTAY-------------- 721

Db        503 RMHRYRAPGEQDGDAMPLGSAVDILATDDPNFSQEDQQDTQIYEKHDNLLHGTKKKKEKM 562

Qy        722 --ITYARQHVH------PTLSDEAAEDLINGYVEMRQKGNFPGSSKKVITATPRQLESMI 773

Db        563 VSAAFMKKYIHVAKIIKPVLTQESATYIAEEYSRLRSQDSMSSDTARTSPVTARTLETLI 622

Qy        774 RISEALARMRFSEVVEKVDAAEAVRLLDVAL---------------------------- 804

Db        623 RLATAHAKARMSKTVDLQDAEEAVELVQYAYFKKVLEKEKKRKKRSEDESETEDEEEKSQ 682
```

```
  Query Match             20.2%;  Score 930;  DB 6;  Length 904;
  Best Local Similarity   28.8%;  Pred. No. 1.1e-57;
  Matches  258;  Conservative  160;  Mismatches  336;  Indels  142;  Gaps   23;

Qy         37 NAASPVAGRRAVRQTPTSAVRRRGRETDSARRRRSRSRSLGNSVYSSPYDAGTPGTPGTP 96
Db          9 SASSPARQRRRISDPLTSSPGR------SSRRADALTSSPGRDL--PPFEDESEGLLGT- 59

Qy         97 VATPVYATPVGTPM---GTPSFHRGTPQ-----------------------------YK 123
Db         60 -EGPMEEEEDGEELIGDGMERDYRPIPELDVYEAEGLALDDEDVEELTASQREAAERTMR 118

Qy        124 QRSELGSQGKPLHRRRRS---QSREPGHRSPSREPSADGRPSESAEPD----------DT 170
Db        119 QRDR--EAGRGLGRMRRGLLYDSSEEDEERPARKRRHVERATEDGEEDEEMIESIENLED 176

Qy        171 LGG----EYAYVWGTNVNIPDVLRAIRRFLHNY-RSSAHDLNSKYIQIIEETVEREEDTL 225
Db        177 LKGHSVRERVSMAGPRL---EIHHRFKNFLRTHVDSHGHNV---FKERISDMCKENRESL 230

Qy        226 NIDMSDIYDHDPDLYAKIVRYPLDIIPLLDTECQEVATSLLPTFEK---HIEARPFNLKA 282
Db        231 VVNYEDLAAREHVLAYFLPEAPAELLQIFDEAALEVVLAMYPKYDRITNHIHVRISHLPL 290

Qy        283 SVHMRELNPSDIDKLVSVKGMVIRCSSIIPEIKGAFFKCLVCGHSPPLVTVVKGRVEEPT 342
Db        291 VEELRSLRQLHLNQLIRTSGVVTSCTGVLPQLSMVKYNCSKCNFVLGPFCQSQNQEVKPG 350

Qy        343 RCEKPECAARNAMSLIHNRCTFANKQIVRLQETPDAIPEGETPHTVSMCLYNTMVDAVKP 402
Db        351 SC--PECQSAGPFEINMEETIYQNYQRIRIQESPGKVAAGRLPRSKDAILLADLVDSCKP 408

Qy        403 GDRIEVTGVFKAMAVRVGPNQRTLRA--LYKTYIDCVHVKKSDR----GRLQTEDPMEMD 456
Db        409 GDEIELTGIYHN---NYDGSLNTANGFPVFATIILANHVAKKDNKVAVGELTDED----- 460

Qy        457 KENDMYAGYHESDTSEAANEAKIQKLKELSKLPGIYDRLSRSLAPSIWELEDIKKGLLCQ 516
Db        461 ---------------------VKMITGLSKDQQIGEKIFASIAPSIYGHEDIKRGLALA 498

Qy        517 LFGGKAKKIPSGASFRGDINVLLVGDPGTSKSQLLQYVHKIAPRGIYTSGRGSSAVGLTA 576
Db        499 LFGGEPKNPGGKHKVRGDINVLLCGDPGTAKSQFLKYIEKVSSRAIFTTGQGASAVGLTA 558

Qy        577 YVTKDPETRETVLESGALVLSDRGICCIDEFDKMSDNARSMLHEVMEQQTVSVAKGGIIA 636
Db        559 YVQRHPVSREWTLEAGALVLADRGVCLIDEFDKMNDQDRTSIHEAMEQQSISISKAGIVT 618

Qy        637 SLNARTSVLACANPSGSRYNARLSVIDNIQLPPTLLSRFDLIYLMLDKPDEQNDRRLARH 696
Db        619 SLQARCTVIAAANPIGGRYDPSLTFSENVDLTEPIISRFDVLCVVRDTVDPVQDEMLARF 678

Qy        697 LVALHYENYEVSKQD--------------------ALDLQTLTAYITYARQHVHPTLSDE 736
Db        679 VVGSHVRHHPSNKKDEGLTNGGTLEPAMPNTYGVEPLPQEVLKKYIIYAKERVRPKLNQM 738

Qy        737 AAEDLINGYVEMRQKGNFPGSSKKVITATPRQLESMIRISEALARMRFSEVVEKVDAAEA 796
Db        739 DQDKVARMYSDLRKESMATGS----IPITVRHIESMIRMAEAHARMHLRDYVMEDDVNMA 794

Qy        797 VRLLDVALQQSATDHATGTIDMDLITTGVSASERIRRAN---LLAALRELIADKIS 849
Db        795 IRV----MMESFIDTQKFSVMRSMRKT-FARYLSFRRDNNDLLLFILKQLVAEQVT 845
```

RESULT 4
US-11-087-227-90
; Sequence 90, Application US/11087227
; Publication No. US20050260566A1
; GENERAL INFORMATION:
; APPLICANT: Fischer, Timothy J.
; APPLICANT: Malinowski, Douglas P.
; APPLICANT: Taylor, Adriann J.
; APPLICANT: Parker, Margaret R.
; TITLE OF INVENTION: METHODS AND COMPOSITIONS FOR THE
; TITLE OF INVENTION: DETECTION OF CERVICAL DISEASE
; FILE REFERENCE: 046143/287139
; CURRENT APPLICATION NUMBER: US/11/087,227
; CURRENT FILING DATE: 2005-03-23
; PRIOR APPLICATION NUMBER: 60/556,495
; PRIOR FILING DATE: 2004-03-24
; NUMBER OF SEQ ID NOS: 90
; SOFTWARE: FastSEQ for Windows Version 4.0
; SEQ ID NO 90
; LENGTH: 821
; TYPE: PRT
; ORGANISM: Homo sapiens
US-11-087-227-90

```
  Query Match             19.9%;  Score 916.5;  DB 7;  Length 821;
  Best Local Similarity   32.3%;  Pred. No. 8.2e-57;
  Matches  252;  Conservative  114;  Mismatches  282;  Indels  133;  Gaps   22;

Qy        195 FLHNYRSSAHDLNSKYIQIIEETVEREEDTLNIDMSDIYDHDPDLYAKIVRYPLDIIPLL 254
Db         33 FLEEFQSS--DGEIKYLQLAEELIRPERNTLVVSFVDLEQFNQQLSTTIQEEFYRVYPYL 90

Qy        255 DTECQEVATSLLPTFEKHIEARPF---------NLKASVHMRELNPSDIDKLVSVKGMVI 305
Db         91 ---CR-----ALKTFVKDRKEIPLAKDFYVAFQDLPTRHKIRELTSSRIGLLTRISGQVV 142

Qy        306 RCSSIIPEIKGAFFKCLVCGHSPPLVTVVKG-----RVEEPTRCEKPECAARNAMSLIHN 360
Db        143 RTHPVHPELVSGTFLCLDC------QTVIRDVEQQFKYTQPNICRNPVCANRRRFLLDTN 196

Qy        361 RCTFANKQIVRLQETPDAIPEGETPHTVSMCLYNTMVDAVKPGDRIEVTGV--------- 411
Db        197 KSRFVDFQKVRIQETQAELPRGSIPRSLEVILRAEAVESAQAGDKCDFTGTLIVVPDVSK 256

Qy        412 FKAMAVRVGPNQRT-------------LRAL------YK-TYIDCVHVKKSDRGRLQTED 451
Db        257 LSTPGARAETNSRVSGVDGYETEGIRGLRALGVRDLSYRLVFLACC-------------- 302

Qy        452 PMEMDKENDMYAGYHESDTSEAANEAKIQ-------KLKELSKLPGIYDRLSRSLAPSIW 504
Db        303 ---VAPTNPRFGGKELRDEEQTAESIKNQMTVKEWEKVFEMSQDKNLYHNLCTSLFPTIH 359

Qy        505 ELEDIKKGLLCQLFGGKAKKIPSGASFRGDINVLLVGDPGTSKSQLLQYVHKIAPRGIYT 564
Db        360 GNDEVKRGVLLMLFGGVPKTTGEGTSLRGDINVCIVGDPSTAKSQFLKHVEEFSPRAVYT 419

Qy        565 SGRGSSAVGLTAYVTKDPETRETVLESGALVLSDRGICCIDEFDKMSDNARSMLHEVMEQ 624
Db        420 SGKASSAAGLTAAVVRDEESHEFVIEAGALMLADNGVCCIDEFDKMDVRDQVAIHEAMEQ 479

Qy        625 QTVSVAKGGIIASLNARTSVLACANPSGSRYNARLSVIDNIQLPPTLLSRFDLIYLMLDK 684
Db        480 QTISITKAGVKATLNARTSILAAANPISGHYDRSKSLKQNINLSAPIMSRFDLFFILVDE 539

Qy        685 PDEQNDRRLARHLVALHYENYEVSKQDALDLQTLTAYITYARQHVHPTLSDEAAEDLING 744
Db        540 CNEVTDYAIARRIVDLH-SRIEESIDRVYSLDDIRRYLLFARQ-FKPKISKESEDFIVEQ 597

Qy        745 YVEMRQKGNFPGSSKKVITATPRQLESMIRISEALARMRFSEVVEKVDAAEAVRLL---- 800
Db        598 YKHLRQRDG-SGVTKSSWRITVRQLESMIRLSEAMARMHCCDEVQPKHVKEAFRLLNKSI 656

Qy        801 ------DVALQQS-----ATDHATGTI----DMDLITTGVSA------------------ 827
Db        657 IRVETPDVNLDQEEEIQMEVDEGAGGINGHADSPAPVNGINGYNEDINQESAPKASLRLG 716

Qy        828 -SERIRRANLLAA-LRELIADKISPGSSSGLKTSQL----LEDIRSQSSVDVSLQDIKNA 881
Db        717 FSEYCRISNLIVLHLRKVEEEE----DESALKRSELVNWYLKEIESEIDSEEELINKKRI 772
```

```
Qy   436 CVHVKKSDRGRLQTEDPMEMDKENDMYAGYHESDTSEAANEAKIQKLKELSKLPGIYDRL 495
Db   300 AHRIVK-------------MNKSEDDESG---------AGELTREELRQIAE-EDFYEKL 336

Qy   496 SRSLAPSIWELEDIKKGLLCQLFGGKAKKIPSGASFRGDINVLLVGDPGTSKSQLLQYVH 555
Db   337 AASIAPEIYGHEDVKKALLLLLVGG-VDQSPRGMKIRGNINICLMGDPGVAKSQLLSYID 395

Qy   556 KIAPRGIYTSGRGSSAVGLTAYVTKDPETRETVLESGALVLSDRGICCIDEFDKMSDNAR 615
Db   396 RLAPRSQYTTGRGSSGVGLTAAVLRDSVSGELTLEGGALVLADQGVCCIDEFDKMAEADR 455

Qy   616 SMLHEVMEQQTVSVAKGGIIASLNARTSVLACANPSGSRYNARLSVIDNIQLPPTLLSRF 675
Db   456 TAIHEVMEQQTISIAKAGILTTLNARCSILAAANPAYGRYNPRRSLEQNIQLPAALLSRF 515

Qy   676 DLIYLMLDKPDEQNDRRLARHLVALHYENYEVSKQ-DALDLQTLTAYITYARQHVHPTLS 734
Db   516 DLLWLIQDRPDRDNDLRLAQHITYVHQHSRQPPSQFEPLDMKLMRRYIAMCREK-QPMVP 574

Qy   735 DEAAEDLINGYVEMRQKGNFPGSSKKVITATPRQLESMIRISEALARMRFSEVVEKVDAA 794
Db   575 ESLADYITAAYVEMRREA---WASKDATYTSARTLLAILRLSTALARLRMVDVVEKEDVN 631

Qy   795 EAVRLLDVALQQSATDHATGTIDMDLITTGVSASERIRRANLLAALRELIA 845
Db   632 EAIRLMEMSKDSLLGD-----------KGQTARTQRPADVIFATVRELVS 670
```

RESULT 2
US-11-087-227-12
; Sequence 12, Application US/11087227
; Publication No. US20050260566A1
; GENERAL INFORMATION:
; APPLICANT: Fischer, Timothy J.
; APPLICANT: Malinowski, Douglas P.
; APPLICANT: Taylor, Adriann J.
; APPLICANT: Parker, Margaret R.
; TITLE OF INVENTION: METHODS AND COMPOSITIONS FOR THE
; TITLE OF INVENTION: DETECTION OF CERVICAL DISEASE
; FILE REFERENCE: 046143/287139
; CURRENT APPLICATION NUMBER: US/11/087,227
; CURRENT FILING DATE: 2005-03-23
; PRIOR APPLICATION NUMBER: 60/556,495
; PRIOR FILING DATE: 2004-03-24
; NUMBER OF SEQ ID NOS: 90
; SOFTWARE: FastSEQ for Windows Version 4.0
; SEQ ID NO 12
; LENGTH: 904
; TYPE: PRT
; ORGANISM: Homo sapiens
US-11-087-227-12

```
  Query Match            20.5%;  Score 943.5;  DB 7;  Length 904;
  Best Local Similarity  29.0%;  Pred. No. 1.2e-58;
  Matches  261;  Conservative  157;  Mismatches  343;  Indels  139;  Gaps   23;

Qy    31 TSPSFDNAASPVAGRRAVRQTPTSAVRRRGRETDSARRRRSRSRSLGNSVYSSPYDAGTP 90
Db     4 SSESFTMASSP-AQRRRGNDPLTSSPGRSSRRTDAL--TSSPGRDL------PPFEDESE 54

Qy    91 GTPGTPVATPVYATPVGTPM---GTPSFHRGTPQYKQRSELG------------SQGKP 134
Db    55 GLLGT--EGPLEEEEDGEELIGDGMERDYRAIPELDAYEAEGLALDDEDVEELTASQREA 112

Qy   135 LHRRRRSQSREPG-----------------HRSPSREPSADGRPSESAEPD-------- 168
Db   113 AERAMRQRDREAGRGLGRMRRGLLYDSDEEDEERPARKRRQVERATEDGEEDEEMIESIE 172

Qy   169 --DTLGGEYAYVW----GTNVNIPDVLRAIRRFLHNY-RSSAHDLNSKYIQIIEETVERE 221
Db   173 NLEDLKGHSVREWVSMAGPRL---EIHHRFKNFLRTHVDSHGHNV---FKERISDMCKEN 226

Qy   222 EDTLNIDMSDIYDHDPDLYAKIVRYPLDIIPLLDTECQEVATSLLPTFEK---HIEARPF 278
Db   227 RESLVVNYEDLAAREHVLAYFLPEAPAELLQIFDEAALEVVLAMYPKYDRITNHIHVRIS 286

Qy   279 NLKASVHMRELNPSDIDKLVSVKGMVIRCSSIIPEIKGAFFKCLVCGHSPPLVTVVKGRV 338
Db   287 HLPLVEELRSLRQLHLNQLIRTSGVVTSCTGVLPQLSMVKYNCNKCNFVLGPFCQSQNQE 346

Qy   339 EEPTRCEKPECAARNAMSLIHNRCTFANKQIVRLQETPDAIPEGETPHTVSMCLYNTMVD 398
Db   347 VKPGSC--PECQSAGPFEVNMEETIYQNYQRIRIQESPGKVAAGRLPRSKDAILLADLVD 404

Qy   399 AVKPGDRIEVTGVFKAMAVRVGPNQRTLRA--LYKTYIDCVHVKKSDR----GRLQTEDP 452
Db   405 SCKPGDEIELTGIYHN---NYDGSLNTANGFPVFATVILANHVAKKDNKVAVGELTDED- 460

Qy   453 MEMDKENDMYAGYHESDTSEAANEAKIQKLKELSKLPGIYDRLSRSLAPSIWELEDIKKG 512
Db   461 --------------------------VKMITSLSKDQQIGEKIFASIAPSIYGHEDIKRG 494

Qy   513 LLCQLFGGKAKKIPSGASFRGDINVLLVGDPGTSKSQLLQYVHKIAPRGIYTSGRGSSAV 572
Db   495 LALALFGGEPKNPGGKHKVRGDINVLLCGDPGTAKSQFLKYIEKVSSRAIFTTGQGASAV 554

Qy   573 GLTAYVTKDPETRETVLESGALVLSDRGICCIDEFDKMSDNARSMLHEVMEQQTVSVAKG 632
Db   555 GLTAYVQRHPVSREWTLEAGALVLADRGVCLIDEFDKMNDQDRTSIHEAMEQQSISISKA 614

Qy   633 GIIASLNARTSVLACANPSGSRYNARLSVIDNIQLPPTLLSRFDLIYLMLDKPDEQNDRR 692
Db   615 GIVTSLQARCTVIAAANPIGGRYDPSLTFSENVDLTEPIISRFDILCVVRDTVDPVQDEM 674

Qy   693 LARHLVALHYENYEVSKQ-------------------DALDLQTLTAYITYARQHVHPT 732
Db   675 LARFVVGSHVRHHPSNKEEEGLANGSAAEPAMPNTYGVEPLPQEVLKKYIIYAKERVHPK 734

Qy   733 LSDEAAEDLINGYVEMRQKGNFPGSSKKVITATPRQLESMIRISEALARMRFSEVVEKVD 792
Db   735 LNQMDQDKVAKMYSDLRKESMATGS----IPITVRHIESMIRMAEAHARIHLRDYVIEDD 790

Qy   793 AAEAVRLLDVALQQSATDHATGTIDMDLITTGVSASERIRRAN---LLAALRELIADKIS 849
Db   791 VNMAIRV----MLESFIDTQKFSVMRSMRKT-FARYLSFRRDNNELLLFILKQLVAEQVT 845
```

RESULT 3
US-10-507-275-5
; Sequence 5, Application US/10507275
; Publication No. US20050250166A1
; GENERAL INFORMATION:
; APPLICANT: Masai, Hisao
; APPLICANT: Tamai, Katsuyuki
; APPLICANT: Medical and Biological Laboratories Co., Ltd.
; APPLICANT: Japan Science and Technology Agency
; APPLICANT: Ginkgo Biomedical Research Institute Co., Ltd.
; TITLE OF INVENTION: Cdc7-ASK Kinase Complex, Substrates of the Kinase Complex,
; TITLE OF INVENTION: Specific Antibodies to the Substrates, and Screening Methods
; TITLE OF INVENTION: Using the Same to Screen for Compounds Comprising Cdc7-ASK
; TITLE OF INVENTION: Kinase Inhititory Ability
; FILE REFERENCE: 082368-001100US
; CURRENT APPLICATION NUMBER: US/10/507,275
; CURRENT FILING DATE: 2004-09-09
; PRIOR APPLICATION NUMBER: JP 2002-067702
; PRIOR FILING DATE: 2002-03-12
; PRIOR APPLICATION NUMBER: WO PCT/JP03/02918
; PRIOR FILING DATE: 2003-03-12
; NUMBER OF SEQ ID NOS: 21
; SOFTWARE: PatentIn Ver. 2.1
; SEQ ID NO 5
; LENGTH: 904
; TYPE: PRT
; ORGANISM: Mus musculus
US-10-507-275-5

```
OM protein - protein search, using sw model

Run on:         December  7, 2005, 17:39:01 ; Search time 12 Seconds
                                   (without alignments)
                                   419.297 Million cell updates/sec

Title:          US-10-768-511-8
Perfect score:  4609
Sequence:       1 MENNDALDIGAVSSPYPSQS..........LGSLQGEGFLTVHGDIVKRV 901

Scoring table:  BLOSUM62
                Gapop 10.0 , Gapext 0.5

Searched:       32527 seqs, 5584426 residues

Total number of hits satisfying chosen parameters:      32527

Minimum DB seq length: 0
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%
                 Maximum Match 100%
                 Listing first 45 summaries

Database :       Published_Applications_AA_New:*
               1:  /cgn2_6/ptodata/2/pubpaa/US09_NEW_PUB.pep:*
               2:  /cgn2_6/ptodata/2/pubpaa/US06_NEW_PUB.pep:*
               3:  /cgn2_6/ptodata/2/pubpaa/US07_NEW_PUB.pep:*
               4:  /cgn2_6/ptodata/2/pubpaa/US08_NEW_PUB.pep:*
               5:  /cgn2_6/ptodata/2/pubpaa/PCT_NEW_PUB.pep:*
               6:  /cgn2_6/ptodata/2/pubpaa/US10_NEW_PUB.pep:*
               7:  /cgn2_6/ptodata/2/pubpaa/US11_NEW_PUB.pep:*
               8:  /cgn2_6/ptodata/2/pubpaa/US60_NEW_PUB.pep:*

     Pred. No. is the number of results predicted by chance to have a
     score greater than or equal to the score of the result being printed,
     and is derived by analysis of the total score distribution.
```

SUMMARIES

Result No.	Score	% Query Match	Length	DB	ID	Description
1	960.5	20.8	719	7	US-11-087-227-14	Sequence 14, Appl
2	943.5	20.5	904	7	US-11-087-227-12	Sequence 12, Appl
3	930	20.2	904	6	US-10-507-275-5	Sequence 5, Appli
4	916.5	19.9	821	7	US-11-087-227-90	Sequence 90, Appl
5	870	18.9	543	7	US-11-087-227-16	Sequence 16, Appl
6	845	18.3	808	6	US-10-878-556A-80	Sequence 80, Appl
7	835	18.1	892	6	US-10-507-275-3	Sequence 3, Appli
8	150	3.3	1493	7	US-11-004-057-4	Sequence 4, Appli
9	145.5	3.2	1493	7	US-11-004-057-21	Sequence 21, Appl
10	129	2.8	420	7	US-11-074-176-48	Sequence 48, Appl
11	129	2.8	498	6	US-10-467-657-7810	Sequence 7810, Ap
12	128.5	2.8	2004	6	US-10-467-657-84	Sequence 84, Appl
13	128.5	2.8	2004	6	US-10-467-657-6322	Sequence 6322, Ap
14	125.5	2.7	1299	6	US-10-821-234-1145	Sequence 1145, Ap
15	122	2.6	824	6	US-10-957-569-31	Sequence 31, Appl
16	120.5	2.6	5024	6	US-10-793-626-2964	Sequence 2964, Ap
17	118	2.6	1249	7	US-11-039-398-22	Sequence 22, Appl
18	118	2.6	1252	7	US-11-039-398-20	Sequence 20, Appl
19	116	2.5	4384	6	US-10-821-234-1120	Sequence 1120, Ap
20	115	2.5	825	7	US-11-074-176-20	Sequence 20, Appl
21	112	2.4	1095	6	US-10-793-626-3154	Sequence 3154, Ap
22	111	2.4	665	6	US-10-793-626-1232	Sequence 1232, Ap
23	111	2.4	1232	7	US-11-039-398-18	Sequence 18, Appl
24	111	2.4	1235	7	US-11-039-398-16	Sequence 16, Appl
25	110	2.4	1304	6	US-10-821-234-1648	Sequence 1648, Ap
26	109.5	2.4	820	6	US-10-467-657-4910	Sequence 4910, Ap
27	109.5	2.4	874	6	US-10-510-386-28	Sequence 28, Appl
28	109.5	2.4	1047	6	US-10-510-386-200	Sequence 200, App
29	109	2.4	896	6	US-10-467-657-7004	Sequence 7004, Ap
30	108	2.3	469	6	US-10-793-626-276	Sequence 276, App
31	107.5	2.3	895	7	US-11-150-406-2	Sequence 2, Appli
32	107.5	2.3	920	6	US-10-821-234-1129	Sequence 1129, Ap
33	105.5	2.3	612	6	US-10-467-657-3988	Sequence 3988, Ap
34	105	2.3	431	7	US-11-058-735-79	Sequence 79, Appl
35	105	2.3	668	6	US-10-467-657-4238	Sequence 4238, Ap
36	105	2.3	1302	7	US-11-004-057-6	Sequence 6, Appli
37	104.5	2.3	274	6	US-10-821-234-956	Sequence 956, App
38	104	2.3	1189	7	US-11-074-176-134	Sequence 134, App
39	104	2.3	1596	7	US-11-060-005-4	Sequence 4, Appli
40	104	2.3	2828	7	US-11-080-991-54	Sequence 54, Appl
41	104	2.3	2828	7	US-11-186-284-49	Sequence 49, Appl
42	103.5	2.2	1210	6	US-10-624-932-26	Sequence 26, Appl
43	103	2.2	695	6	US-10-467-962B-47	Sequence 47, Appl
44	102.5	2.2	1933	6	US-10-523-912-2	Sequence 2, Appli
45	102	2.2	1766	7	US-11-075-185-10	Sequence 10, Appl

ALIGNMENTS

```
RESULT 1
US-11-087-227-14
; Sequence 14, Application US/11087227
; Publication No. US20050260566A1
; GENERAL INFORMATION:
;  APPLICANT: Fischer, Timothy J.
;  APPLICANT:  Malinowski, Douglas P.
;  APPLICANT:  Taylor, Adriann J.
;  APPLICANT:  Parker, Margaret R.
;  TITLE OF INVENTION: METHODS AND COMPOSITIONS FOR THE
;  TITLE OF INVENTION:  DETECTION OF CERVICAL DISEASE
;  FILE REFERENCE: 046143/287139
;  CURRENT APPLICATION NUMBER: US/11/087,227
;  CURRENT FILING DATE:  2005-03-23
;  PRIOR APPLICATION NUMBER: 60/556,495
;  PRIOR FILING DATE: 2004-03-24
;  NUMBER OF SEQ ID NOS: 90
;  SOFTWARE: FastSEQ for Windows Version 4.0
; SEQ ID NO 14
;   LENGTH: 719
;   TYPE: PRT
;   ORGANISM: Homo sapiens
US-11-087-227-14

  Query Match             20.8%;  Score 960.5;  DB 7;  Length 719;
  Best Local Similarity   33.2%;  Pred. No. 5.6e-60;
  Matches  236;  Conservative  137;  Mismatches  228;  Indels  110;  Gaps   21;

Qy        192 IRRFLHNYRSSAHDLNSKYIQIIEETV---EREEDTLNIDMSDIYDHDPDL--------- 239
              :::||  :      :|  |  :   :  |     ||:  | :|: |: :  ||:|
Db         13 VKKFLQEFYQD-DELGKKQFKYGNQLVRLAHREQVALYVDLDDVAEDDPELVDSICENAR 71

Qy        240 -YAKIVRYPL-DIIP-----------LLDTECQ--------------------EVATSLL 266
               |||:    : :::|           :||   :                    :    |:
Db         72 RYAKLFADAVQELLPQYKEREVVNKDVLDVYIEHRLMMEQRSRDPGMVRSPQNQYPAELM 131

Qy        267 PTFEKHIEARPFNLKASVHMRELNPSDIDKLVSVKGMVIRCSSIIPEIKGAFFKCLVCG- 325
                || : :   | : |  | :||:    : |||:|:|:|  |  |  : |::  | :  |  ||
Db        132 RRFELYFQG-PSSNKPRV-IREVRADSVGKLVTVRGIVTRVSEVKPKMVVATYTCDQCGA 189

Qy        326 --HSPPLVTVVKGRVEEPT-----RCEKPECAARNA---MSLIHNRCTFANKQIVRLQET 375
                : |           :: ||      |   ||   :   : |       |   | :::||
Db        190 ETYQP---------IQSPTFMPLIMCPSQECQTNRSGGRLYLQTRGSRFIKFQEMKMQEH 240

Qy        376 PDAIPEGETPHTVSMCLYNTMVDAVKPGDRIEVTGVFKAMAVRVGPNQRTLRALYKTYID 435
                | :| |  | :::: :              :||| : |||:|  : :| | |      | :||::
Db        241 SDQVPVGNIPRSITVLVEGENTRIAQPGDHVSVTGIFLPI-LRTGFRQVVQGLLSETYLE 299
```

```
OM protein - nucleic search, using frame_plus_p2n model

Run on:         December  6, 2005, 09:52:39 ; Search time 280 Seconds
                                          (without alignments)
                                          5719.939 Million cell updates/sec

Title:          US-10-768-511-8
Perfect score:  4609
Sequence:       1 MENNDALDIGAVSSPYPSQS..........LGSLQGEGFLTVHGDIVKRV 901

Scoring table:  BLOSUM62
                Xgapop 10.0 , Xgapext  0.5
                Ygapop 10.0 , Ygapext  0.5
                Fgapop  6.0 , Fgapext  7.0
                Delop   6.0 , Delext   7.0

Searched:       1303057 seqs, 888780828 residues

Total number of hits satisfying chosen parameters:      2606114

Minimum DB seq length: 0
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%
                 Maximum Match 100%
                 Listing first 45 summaries

Command line parameters:
-MODEL=frame+_p2n.model -DEV=xlh
-Q=/cgn2_1/USPTO_spool/US10768511/runat_02122005_102335_18440/app_query.fasta_1.1095
-DB=Issued_Patents_NA -QFMT=fastap -SUFFIX=rni -MINMATCH=0.1 -LOOPCL=0
-LOOPEXT=0 -UNITS=bits -START=1 -END=-1 -MATRIX=blosum62 -TRANS=human40.cdi
-LIST=45 -DOCALIGN=200 -THR_SCORE=pct -THR_MAX=100 -THR_MIN=0 -ALIGN=15
-MODE=LOCAL -OUTFMT=pto -NORM=ext -HEAPSIZE=500 -MINLEN=0 -MAXLEN=2000000000
-USER=US10768511_@CGN_1_1_162_@runat_02122005_102335_18440 -NCPU=6 -ICPU=3
-NO_MMAP -LARGEQUERY -NEG_SCORES=0 -WAIT -DSPBLOCK=100 -LONGLOG
-DEV_TIMEOUT=120 -WARN_TIMEOUT=30 -THREADS=1 -XGAPOP=10 -XGAPEXT=0.5 -FGAPOP=6
-FGAPEXT=7 -YGAPOP=10 -YGAPEXT=0.5 -DELOP=6 -DELEXT=7

Database :      Issued_Patents_NA:*
                1:  /cgn2_6/ptodata/1/ina/1_COMB.seq:*
                2:  /cgn2_6/ptodata/1/ina/5_COMB.seq:*
                3:  /cgn2_6/ptodata/1/ina/6A_COMB.seq:*
                4:  /cgn2_6/ptodata/1/ina/6B_COMB.seq:*
                5:  /cgn2_6/ptodata/1/ina/H_COMB.seq:*
                6:  /cgn2_6/ptodata/1/ina/PCTUS_COMB.seq:*
                7:  /cgn2_6/ptodata/1/ina/PP_COMB.seq:*
                8:  /cgn2_6/ptodata/1/ina/RE_COMB.seq:*
                9:  /cgn2_6/ptodata/1/ina/backfiles1.seq:*
```

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

	Result No.	Score	% Query Match	Length	DB	ID	Description
	1	4609	100.0	4348	3	US-09-828-062-5	Sequence 5, Appli
	2	1139	24.7	723	3	US-09-828-062-2	Sequence 2, Appli
	3	993	21.5	2533	3	US-09-604-605-1	Sequence 1, Appli
	4	993	21.5	2533	3	US-10-101-080-1	Sequence 1, Appli
	5	982.5	21.3	3000	3	US-09-949-016-2621	Sequence 2621, Ap
	6	946.5	20.5	3378	3	US-09-964-899-48	Sequence 48, Appl
	7	939	20.4	2533	3	US-09-949-016-5021	Sequence 5021, Ap
	8	936	20.3	3445	3	US-09-976-594-323	Sequence 323, App
	9	918.5	19.9	2917	3	US-09-949-016-2468	Sequence 2468, Ap
	10	856	18.6	3379	3	US-09-220-132-12	Sequence 12, Appl
	11	817.5	17.7	1635	3	US-09-248-796A-4534	Sequence 4534, Ap
	12	734.5	15.9	1488	3	US-09-248-796A-5904	Sequence 5904, Ap
	13	670	14.5	996	3	US-09-248-796A-4570	Sequence 4570, Ap
c	14	656	14.2	1664976	3	US-08-916-421B-1	Sequence 1, Appli
c	15	656	14.2	1664976	3	US-09-692-570-1	Sequence 1, Appli
	16	474	10.3	1305	3	US-09-248-796A-4711	Sequence 4711, Ap
	17	438.5	9.5	507	3	US-09-401-064-158	Sequence 158, App
	18	427	9.3	12932	3	US-09-949-016-14363	Sequence 14363, A
	19	391.5	8.5	519	3	US-09-248-796A-5916	Sequence 5916, Ap
	20	381.5	8.3	1290	3	US-09-248-796A-4988	Sequence 4988, Ap
	21	339	7.4	311	3	US-09-313-294A-83	Sequence 83, Appl
	22	333	7.2	294	3	US-09-313-294A-4610	Sequence 4610, Ap
	23	330.5	7.2	58407	3	US-08-916-421B-2	Sequence 2, Appli
	24	330.5	7.2	58407	3	US-09-692-570-2	Sequence 2, Appli
	25	312	6.8	297	3	US-09-313-294A-4978	Sequence 4978, Ap
	26	289	6.3	2313	3	US-09-799-451-937	Sequence 937, App
	27	266.5	5.8	613	3	US-09-533-559-658	Sequence 658, App
	28	246.5	5.3	492	3	US-09-248-796A-5915	Sequence 5915, Ap
	29	235	5.1	28366	3	US-09-949-016-16763	Sequence 16763, A
	30	227	4.9	273	3	US-09-313-294A-1684	Sequence 1684, Ap
	31	222.5	4.8	256	3	US-09-313-294A-2519	Sequence 2519, Ap
	32	219	4.8	747	3	US-09-248-796A-4987	Sequence 4987, Ap
c	33	201.5	4.4	14636	3	US-09-173-914-6	Sequence 6, Appli
	34	199	4.3	1059	3	US-09-248-796A-5905	Sequence 5905, Ap
c	35	195	4.2	13256	3	US-09-902-540-1006	Sequence 1006, Ap
c	36	189.5	4.1	4403765	3	US-09-103-840A-2	Sequence 2, Appli
c	37	188	4.1	565	3	US-09-328-111-454	Sequence 454, App
	38	185.5	4.0	289	3	US-09-313-294A-3770	Sequence 3770, Ap
	39	175.5	3.8	46899	2	US-08-471-119A-1	Sequence 1, Appli
	40	169.5	3.7	28958	2	US-08-258-261B-6	Sequence 6, Appli
	41	169.5	3.7	28958	2	US-08-456-837-6	Sequence 6, Appli
	42	169.5	3.7	28958	2	US-08-457-342-6	Sequence 6, Appli
	43	169.5	3.7	28958	2	US-08-457-646A-6	Sequence 6, Appli
	44	169.5	3.7	28958	2	US-08-458-076A-6	Sequence 6, Appli
	45	169.5	3.7	28958	2	US-08-764-233A-4	Sequence 4, Appli

ALIGNMENTS

```
RESULT 1
US-09-828-062-5
; Sequence 5, Application US/09828062
; Patent No. 6710229
; GENERAL INFORMATION:
;  APPLICANT: COSTA E SILVA, OSWALDO DA
;  APPLICANT:  BOHNERT, HANS J.
;  APPLICANT:  VAN THIELEN, NOCHA
;  APPLICANT:  CHEN, ROUYING
;  APPLICANT:  SARRIA-MILLAN, RODRIGO
;  TITLE OF INVENTION: CELL CYCLE STRESS-RELATED PROTEINS AND METHODS OF USE
;  TITLE OF INVENTION:  IN PLANTS
;  FILE REFERENCE: 16313-0031
;  CURRENT APPLICATION NUMBER: US/09/828,062
;  CURRENT FILING DATE:  2001-08-20
;  PRIOR APPLICATION NUMBER: 60/196,001
;  PRIOR FILING DATE: 2000-04-07
;  NUMBER OF SEQ ID NOS: 34
;  SOFTWARE: PatentIn Ver. 2.1
; SEQ ID NO 5
;   LENGTH: 4348
;   TYPE: DNA
;   ORGANISM: Physcomitrella patens
US-09-828-062-5

Alignment Scores:
Pred. No.:               0              Length:        4348
Score:                   4609.00        Matches:       901
Percent Similarity:      100.00%        Conservative:  0
Best Local Similarity:   100.00%        Mismatches:    0
Query Match:             100.00%        Indels:        0
DB:                      3              Gaps:          0
```

```
US-10-768-511-8 (1-901) x US-09-828-062-5 (1-4348)

Qy       1 MetGluAsnAsnAspAlaLeuAspIleGlyAlaValSerSerProTyrProSerGlnSer 20
Db    1554 ATGGAAAATAATGATGCACTTGACATTGGAGCCGTGTCGTCCCCATATCCTTCGCAATCT 1613

Qy      21 GluGlyValSerThrProLeuProGlnValThrSerProSerPheAspAsnAlaAlaSer 40
Db    1614 GAAGGAGTGTCTACGCCATTGCCGCAAGTAACATCACCGAGCTTCGACAATGCAGCCTCA 1673

Qy      41 ProValAlaGlyArgArgAlaValArgGlnThrProThrSerAlaValArgArgArgGly 60
Db    1674 CCCGTGGCCGGGCGGAGGGCCGTACGGCAGACCCCTACATCTGCAGTTCGAAGGAGAGGG 1733

Qy      61 ArgGluThrAspSerAlaArgArgArgArgSerArgSerArgSerLeuGlyAsnSerVal 80
Db    1734 AGAGAAACGGATTCCGCTCGTCGTAGGAGGAGTCGATCTCGCAGTTTAGGCAATTCTGTT 1793

Qy      81 TyrSerSerProTyrAspAlaGlyThrProGlyThrProGlyThrProValAlaThrPro 100
Db    1794 TATAGTTCCCCTTACGATGCGGGGACTCCTGGAACTCCTGGAACTCCAGTGGCTACTCCG 1853

Qy     101 ValTyrAlaThrProValGlyThrProMetGlyThrProSerPheHisArgGlyThrPro 120
Db    1854 GTTTACGCTACCCCAGTCGGTACACCTATGGGTACCCCATCGTTCCATCGTGGCACGCCA 1913

Qy     121 GlnTyrLysGlnArgSerGluLeuGlySerGlnGlyLysProLeuHisArgArgArgArg 140
Db    1914 CAGTACAAACAGCGCAGTGAGCTTGGTTCCCAGGGGAAGCCTCTACATCGGAGACGTCGA 1973

Qy     141 SerGlnSerArgGluProGlyHisArgSerProSerArgGluProSerAlaAspGlyArg 160
Db    1974 TCTCAATCCAGAGAACCCGGGCATCGATCTCCTTCAAGGGAACCTAGTGCTGATGGGCGT 2033

Qy     161 ProSerGluSerAlaGluProAspAspThrLeuGlyGlyGluTyrAlaTyrValTrpGly 180
Db    2034 CCCTCTGAATCTGCTGAGCCAGATGACACTTTGGGTGGAGAATATGCTTATGTTTGGGGG 2093

Qy     181 ThrAsnValAsnIleProAspValLeuArgAlaIleArgArgPheLeuHisAsnTyrArg 200
Db    2094 ACGAATGTTAACATTCCAGATGTGCTTAGGGCGATTCGTCGATTTCTCCACAATTATCGT 2153

Qy     201 SerSerAlaHisAspLeuAsnSerLysTyrIleGlnIleIleGluGluThrValGluArg 220
Db    2154 TCGAGTGCTCATGATCTTAATTCCAAGTACATCCAGATCATAGAGGAGACTGTGGAGCGT 2213

Qy     221 GluGluAspThrLeuAsnIleAspMetSerAspIleTyrAspHisAspProAspLeuTyr 240
Db    2214 GAGGAGGATACTCTAAATATCGACATGTCAGACATTTATGACCATGATCCTGATCTATAC 2273

Qy     241 AlaLysIleValArgTyrProLeuAspIleIleProLeuLeuAspThrGluCysGlnGlu 260
Db    2274 GCAAAAATTGTTCGATACCCACTCGACATCATCCCCCTGTTGGACACTGAGTGTCAGGAA 2333

Qy     261 ValAlaThrSerLeuLeuProThrPheGluLysHisIleGluAlaArgProPheAsnLeu 280
Db    2334 GTTGCTACCTCTTTACTACCAACGTTTGAGAAGCATATTGAGGCCAGACCTTTCAATCTC 2393

Qy     281 LysAlaSerValHisMetArgGluLeuAsnProSerAspIleAspLysLeuValSerVal 300
Db    2394 AAAGCATCGGTGCACATGCGTGAACTCAACCCTTCAGATATAGACAAATTGGTTTCTGTT 2453

Qy     301 LysGlyMetValIleArgCysSerSerIleIleProGluIleLysGlyAlaPhePheLys 320
Db    2454 AAAGGAATGGTTATCCGGTGCAGTTCTATCATACCTGAAATTAAGGGGGCCTTCTTCAAA 2513

Qy     321 CysLeuValCysGlyHisSerProProLeuValThrValValLysGlyArgValGluGlu 340
Db    2514 TGTTTAGTGTGTGGTCACTCGCCTCCGCTAGTTACAGTTGTTAAAGGGCGGGTTGAGGAG 2573

Qy     341 ProThrArgCysGluLysProGluCysAlaAlaArgAsnAlaMetSerLeuIleHisAsn 360
Db    2574 CCAACAAGGTGTGAAAAGCCAGAATGTGCAGCACGGAATGCTATGTCTCTTATTCACAAT 2633

Qy     361 ArgCysThrPheAlaAsnLysGlnIleValArgLeuGlnGluThrProAspAlaIlePro 380
Db    2634 CGATGCACTTTTGCAAATAAGCAGATAGTGCGTCTTCAAGAAACTCCAGATGCCATTCCT 2693

Qy     381 GluGlyGluThrProHisThrValSerMetCysLeuTyrAsnThrMetValAspAlaVal 400
Db    2694 GAAGGAGAGACTCCACACACAGTCAGCATGTGTTTATACAACACTATGGTTGATGCTGTG 2753

Qy     401 LysProGlyAspArgIleGluValThrGlyValPheLysAlaMetAlaValArgValGly 420
Db    2754 AAGCCTGGAGATCGTATTGAGGTAACAGGAGTTTTCAAGGCCATGGCAGTTCGAGTTGGT 2813

Qy     421 ProAsnGlnArgThrLeuArgAlaLeuTyrLysThrTyrIleAspCysValHisValLys 440
Db    2814 CCGAATCAACGAACATTACGAGCATTGTATAAGACCTACATCGATTGCGTGCACGTCAAG 2873

Qy     441 LysSerAspArgGlyArgLeuGlnThrGluAspProMetGluMetAspLysGluAsnAsp 460
Db    2874 AAGTCTGACAGGGGTCGACTGCAAACTGAAGATCCTATGGAGATGGATAAGGAGAATGAT 2933

Qy     461 MetTyrAlaGlyTyrHisGluSerAspThrSerGluAlaAlaAsnGluAlaLysIleGln 480
Db    2934 ATGTATGCTGGGTATCATGAAAGTGATACTTCAGAAGCTGCTAATGAAGCAAAGATTCAA 2993

Qy     481 LysLeuLysGluLeuSerLysLeuProGlyIleTyrAspArgLeuSerArgSerLeuAla 500
Db    2994 AAACTTAAAGAGCTGTCCAAGCTCCCGGGCATTTATGATAGACTTTCAAGGTCGCTGGCT 3053

Qy     501 ProSerIleTrpGluLeuGluAspIleLysLysGlyLeuLeuCysGlnLeuPheGlyGly 520
Db    3054 CCAAGCATTTGGGAGCTTGAAGATATTAAAAAGGGTCTTCTTTGCCAGCTCTTTGGTGGG 3113

Qy     521 LysAlaLysLysIleProSerGlyAlaSerPheArgGlyAspIleAsnValLeuLeuVal 540
Db    3114 AAGGCTAAGAAAATTCCATCTGGAGCATCTTTCCGAGGTGACATCAATGTTTTACTTGTT 3173

Qy     541 GlyAspProGlyThrSerLysSerGlnLeuLeuGlnTyrValHisLysIleAlaProArg 560
Db    3174 GGGGACCCTGGTACCAGTAAATCTCAGCTGCTTCAGTATGTGCACAAGATAGCTCCTCGT 3233

Qy     561 GlyIleTyrThrSerGlyArgGlySerSerAlaValGlyLeuThrAlaTyrValThrLys 580
Db    3234 GGAATCTACACTAGTGGGCGAGGAAGTTCGGCGGTTGGGCTGACAGCGTATGTAACGAAG 3293

Qy     581 AspProGluThrArgGluThrValLeuGluSerGlyAlaLeuValLeuSerAspArgGly 600
Db    3294 GATCCAGAAACTCGAGAGACGGTATTGGAGAGCGGAGCTTTGGTTCTTAGTGATCGTGGG 3353

Qy     601 IleCysCysIleAspGluPheAspLysMetSerAspAsnAlaArgSerMetLeuHisGlu 620
Db    3354 ATATGCTGTATCGATGAGTTCGACAAAATGTCTGATAATGCCCGAAGCATGCTTCATGAG 3413

Qy     621 ValMetGluGlnGlnThrValSerValAlaLysGlyGlyIleIleAlaSerLeuAsnAla 640
Db    3414 GTAATGGAGCAACAAACGGTATCTGTAGCCAAAGGGGGTATCATTGCCTCGCTGAACGCT 3473

Qy     641 ArgThrSerValLeuAlaCysAlaAsnProSerGlySerArgTyrAsnAlaArgLeuSer 660
Db    3474 CGGACGTCTGTCCTTGCATGTGCAAATCCTAGTGGGTCCCGATACAATGCGCGCCTTTCT 3533

Qy     661 ValIleAspAsnIleGlnLeuProProThrLeuLeuSerArgPheAspLeuIleTyrLeu 680
Db    3534 GTGATTGATAACATCCAGCTTCCTCCAACTCTACTTTCTAGATTTGATTTAATTTACTTA 3593

Qy     681 MetLeuAspLysProAspGluGlnAsnAspArgArgLeuAlaArgHisLeuValAlaLeu 700
Db    3594 ATGCTCGACAAACCAGACGAGCAAAACGATCGTCGTCTCGCCAGGCATCTCGTGGCTTTA 3653

Qy     701 HisTyrGluAsnTyrGluValSerLysGlnAspAlaLeuAspLeuGlnThrLeuThrAla 720
Db    3654 CACTATGAAAACTATGAAGTTTCAAAGCAGGACGCCTTAGATCTACAAACACTTACCGCG 3713
```

```
Qy        721 TyrIleThrTyrAlaArgGlnHisValHisProThrLeuSerAspGluAlaAlaGluAsp 740
Db       3714 TATATCACCTATGCTCGTCAGCATGTACATCCTACATTAAGTGATGAAGCTGCTGAAGAT 3773

Qy        741 LeuIleAsnGlyTyrValGluMetArgGlnLysGlyAsnPheProGlySerSerLysLys 760
Db       3774 TTGATTAATGGCTATGTTGAGATGCGCCAAAAGGGCAACTTTCCTGGAAGCAGTAAAAAG 3833

Qy        761 ValIleThrAlaThrProArgGlnLeuGluSerMetIleArgIleSerGluAlaLeuAla 780
Db       3834 GTGATAACAGCCACACCTCGGCAACTCGAAAGTATGATTCGTATCAGTGAAGCCCTAGCT 3893

Qy        781 ArgMetArgPheSerGluValValGluLysValAspAlaAlaGluAlaValArgLeuLeu 800
Db       3894 CGAATGAGATTTTCTGAAGTGGTAGAGAAAGTTGATGCAGCAGAAGCTGTGCGCCTTTTA 3953

Qy        801 AspValAlaLeuGlnGlnSerAlaThrAspHisAlaThrGlyThrIleAspMetAspLeu 820
Db       3954 GACGTCGCTTTGCAGCAATCTGCTACTGATCATGCAACAGGTACGATAGACATGGATCTT 4013

Qy        821 IleThrThrGlyValSerAlaSerGluArgIleArgArgAlaAsnLeuLeuAlaAlaLeu 840
Db       4014 ATCACGACTGGAGTGTCGGCCAGCGAGCGTATTCGTCGGGCCAACTTGCTAGCTGCTCTG 4073

Qy        841 ArgGluLeuIleAlaAspLysIleSerProGlySerSerSerGlyLeuLysThrSerGln 860
Db       4074 CGAGAGCTTATAGCAGATAAAATTTCACCTGGCAGCTCCTCTGGCTTGAAGACCAGTCAG 4133

Qy        861 LeuLeuGluAspIleArgSerGlnSerSerValAspValSerLeuGlnAspIleLysAsn 880
Db       4134 CTTCTTGAGGATATCCGGAGCCAAAGCAGTGTGGACGTTAGTTTGCAGGATATTAAAAAT 4193

Qy        881 AlaLeuGlySerLeuGlnGlyGluGlyPheLeuThrValHisGlyAspIleValLysArg 900
Db       4194 GCTCTGGGTAGCCTCCAAGGAGAAGGCTTTCTTACTGTCCATGGTGACATAGTCAAGAGA 4253

Qy        901 Val 901
Db       4254 GTT 4256
```

RESULT 2
US-09-828-062-2
; Sequence 2, Application US/09828062
; Patent No. 6710229
; GENERAL INFORMATION:
; APPLICANT: COSTA E SILVA, OSWALDO DA
; APPLICANT: BOHNERT, HANS J.
; APPLICANT: VAN THIELEN, NOCHA
; APPLICANT: CHEN, ROUYING
; APPLICANT: SARRIA-MILLAN, RODRIGO
; TITLE OF INVENTION: CELL CYCLE STRESS-RELATED PROTEINS AND METHODS OF USE
; TITLE OF INVENTION: IN PLANTS
; FILE REFERENCE: 16313-0031
; CURRENT APPLICATION NUMBER: US/09/828,062
; CURRENT FILING DATE: 2001-08-20
; PRIOR APPLICATION NUMBER: 60/196,001
; PRIOR FILING DATE: 2000-04-07
; NUMBER OF SEQ ID NOS: 34
; SOFTWARE: PatentIn Ver. 2.1
; SEQ ID NO 2
; LENGTH: 723
; TYPE: DNA
; ORGANISM: Physcomitrella patens
US-09-828-062-2

Alignment Scores:

Pred. No.:	5.74e-87	Length:	723
Score:	1139.00	Matches:	226
Percent Similarity:	98.26%	Conservative:	0
Best Local Similarity:	98.26%	Mismatches:	3
Query Match:	24.71%	Indels:	1
DB:	3	Gaps:	0

US-10-768-511-8 (1-901) x US-09-828-062-2 (1-723)

```
Qy        403 GlyAspArgIleGluValThrGlyValPheLysAlaMetAlaValArgValGlyProAsn 422
Db         10 GGAGATCGTATTGAGGTAACAGGAGTTTTCAAGGCCATGGCAGTTCGAGTTGGTCCGAAT 69

Qy        423 GlnArgThrLeuArgAlaLeuTyrLysThrTyrIleAspCysValHisValLysLysSer 442
Db         70 CAACGAACATTACGAGCATTGTATAAGACCTACATCGATTGCGTGCACGTCAAGAAGTCT 129

Qy        443 AspArgGlyArgLeuGlnThrGluAspProMetGluMetAspLysGluAsnAspMetTyr 462
Db        130 GACAGGGGTCGACTGCAAACTGAAGATCCTATGGAGATGGATAAGGAGAATGATATGTAT 189

Qy        463 AlaGlyTyrHisGluSerAspThrSerGluAlaAlaAsnGluAlaLysIleGlnLysLeu 482
Db        190 GCTGGGTATCATGAAAGTGATACTTCAGAAGCTGCTAATGAAGCAAAGATTCAAAAACTT 249

Qy        483 LysGluLeuSerLysLeuProGlyIleTyrAspArgLeuSerArgSerLeuAlaProSer 502
Db        250 AAAGAGCTGTCCAAGCTCCCGGACATTTATGATAGACTTTCAAGGTCGCTGGCTCCAAGC 309

Qy        503 IleTrpGluLeuGluAspIleLysLysGlyLeuLeuCysGlnLeuPheGlyGlyLysAla 522
Db        310 ATTTGGGAGCTTGAAGATATTAAAAAGGGTCTTCTTTGCCAGCTCTTTGGTGGAAAGGCT 369

Qy        523 LysLysIleProSerGlyAlaSerPheArgGlyAspIleAsnValLeuLeuValGlyAsp 542
Db        370 AAGAAAATTCCATCTGGAGCATCTTTCCGAGGTGACATCAATGTTTTACTTGTTGGGGAC 429

Qy        543 ProGlyThrSerLysSerGlnLeuLeuGlnTyrValHisLysIleAlaProArgGlyIle 562
Db        430 CCTGGTACCAGTAAATCTCAGCTGCTTCAGTATGTGCACAAGATAGCTCCTCGTGGAATC 489

Qy        563 TyrThrSerGlyArgGlySerSerAlaValGlyLeuThrAlaTyrVal-ThrLysAspPr 582
Db        490 TACACTAGTGGGCGAGGAAGTTCGGCGGTTGGGCTGACAGCGTATGTAAACGAAGGATCC 549

Qy        582 oGluThrArgGluThrValLeuGluSerGlyAlaLeuValLeuSerAspArgGlyIleCy 602
Db        550 AGAAACTCGAGAGACGGTATTGGAGAGCGGAGCTTTGGTTCTTAGTGATCGTGGGATATG 609

Qy        602 sCysIleAspGluPheAspLysMetSerAspAsnAlaArgSerMetLeuHisGluValMe 622
Db        610 CTGTATCGATGAGTTCGACAAAATGTCTGATAATGCCCGAAGCATGCTTCATGAGGTAAT 669

Qy        622 tGluGlnGlnThrValSerValAlaLys 631
Db        670 GGAGCAACAAACGGTATCTGGACCCAAG 697
```

RESULT 3
US-09-604-605-1
; Sequence 1, Application US/09604605
; Patent No. 6421613
; GENERAL INFORMATION:
; APPLICANT: Ramgopal Nadimpalli
; APPLICANT: Carl R. Simmons
; TITLE OF INVENTION: Maize Prolifera Gene and Uses Thereof
; FILE REFERENCE: 1138
; CURRENT APPLICATION NUMBER: US/09/604,605
; CURRENT FILING DATE: 2000-06-27
; PRIOR APPLICATION NUMBER: 60/143,222
; PRIOR FILING DATE: 1999-07-09
; NUMBER OF SEQ ID NOS: 2
; SOFTWARE: FastSEQ for Windows Version 3.0
; SEQ ID NO 1
; LENGTH: 2533
; TYPE: DNA
; ORGANISM: Zea mays
; FEATURE:
; NAME/KEY: CDS
; LOCATION: (110)...(2269)
US-09-604-605-1

```
Alignment Scores:
Pred. No.:               9.8e-74     Length:         2533
Score:                   993.00      Matches:        281
Percent Similarity:      49.88%      Conservative:   131
Best Local Similarity:   34.02%      Mismatches:     292
Query Match:             21.54%      Indels:         122
DB:                      3           Gaps:           25

US-10-768-511-8 (1-901) x US-09-604-605-1 (1-2533)

Qy     84 ProTyrAspAlaGlyThrProGlyThrProGlyThrProValAlaThrProValTyrAla 103
Db      9 CCTTCTTCCCGCCATCCCCCGAACACTTCGCCGCCCCCTCCCGCCTCCCGCTGTTCTCT 68

Qy    104 Thr-ProValGlyThrProMetGlyThrProSerPheHisArgGlyThrProGlnTyrLy 123
Db     69 GCCTCCCACCCCCACGCCG---------CCGTCCCATCACCGCCGCTGGGATGAAGAACC 119

Qy    123 s------GlnArgSerGluLeuGlySerGlnGlyLysProLeuHisArgArgArgArgSe 141
Db    120 CTGACTTCGCCGCAGACAAAGCGCTCGCCAAGGACTTCCTGTCCAATTTCACCGG----- 174

Qy    141 rGlnSerArgGluProGlyHisArgSerProSerArgGluProSerAlaAspGlyArgPr 161
Db    175 ----------------------------CCCTCACGGCGAGCCCAAGTACCTGAACATCC 206

Qy    161 oSerGluSerAlaGluProAspAspThrLeuGlyGlyGluTyrAlaTyrValTrpGlyTh 181
Db    207 TGCAAGATGTCGCGAACAGGAAGATACGCGCGGTCCAGATCGAGCT---------GGACG 257

Qy    181 rAsnValAsnIleProAspValLeuArgAlaIleArgArgPheLeuHisAsnTyrArgSe 201
Db    258 ACCTATTCCACTATAAGGATGTCGACGAGGAGTTCTTGCAGCGTGTCACCGAGAACACGC 317

Qy    201 rSerAlaHis---AspLeuAsnSerLysTyr-IleGlnIleIleGluGlu---------- 216
Db    318 GGCGCTACATCGGCATCTTCGCGGAGGCCATGGACGAGATCATGCCGGAGCCCACAGAGG 377

Qy    217 -----ThrValGluArgGluGluAspThrLeuAsnIleAspMetSerAspIleTyrAspH 235
Db    378 CGTACACCGTCGACGAGGACCAGGACATCCTG---------ATGACGCAGCGTGTGGATG 428

Qy    235 isAspProAspLeuTyrAlaLysIleValArgTyrProLeuAspIleIleProLeuLeuA 255
Db    429 AGGGGGCCGACGGCGGCGCGGAC------------------GGCACGGACCCGCTA---- 466

Qy    255 spThrGluCysGlnGluValAlaThrSerLeuLeuProThrPheGluLysHisIleGluA 275
Db    467 -----------CAGAAAATGCCACCAGAGATCAAACGGTTCTTTGAGGTTTACATCAAGG 515

Qy    275 laArgProPheAsnLeuLysAlaSerValHisMetArgGluLeuAsnProSerAspIleA 295
Db    516 CA------TTCTCAAAGGTGACTCCACTCACCATTCGGCAAGTGAAGGCATCCAACATTG 569

Qy    295 spLysLeuValSerValLysGlyMetValIleArgCysSerSerIleIleProGluIleL 315
Db    570 GTCAGCTTGTGAAAATATCTGGGATTGTTACTCGCTGCTCGGATGTGAAGCCCTTGATGC 629

Qy    315 ysGlyAlaPhePheLysCysLeuValCysGlyHisSerProProLeuValThrValValL 335
Db    630 AGGTTGCTGTTTATACATGTGAAGAATGTGGCTTCGAG------ATATACCAGGAAGTGA 683

Qy    335 ysGlyArgValGluGluPro---ThrArgCysGluLysProGluCys---------AlaA 351
Db    684 CTGCTAGAGTGTTTATGCCCCTCATTGAATGTCCATCTCAACGATGCAAGCTGAACAAAG 743

Qy    351 laArgAsnAlaMetSerLeuIleHisAsnArgCysThrPheAlaAsnLysGlnIleValA 371
Db    744 CGAAGGGAAATCTAATCCTTCAACTGCGTGCGTCAAAGTTTTTGAAATTTCAGGAGGTAA 803

Qy    371 rgLeuGlnGluThrProAspAlaIleProGluGlyGluThrProHisThrValSerMetC 391
Db    804 AGCTTCAAGAGTTAGCAGAGCACGTACCAAAGGGCCACATTCCTCGTGCTCTGACTGTTC 863

Qy    391 ysLeuTyrAsnThrMetValAspAlaValLysProGlyAspArgIleGluValThrGlyV 411
Db    864 ATCTAAGAGGAGAGCTGACTAGAAAGGTTGCACCTGGAGATGTGGTTGAGATGTCGGGTA 923

Qy    411 alPheLysAlaMet------------------AlaValArgValGlyProAsnGlnArgT 425
Db    924 TTTTTCTCCCTATGCCATACTACGGGTTTAGAGCGATGCGCGCAGGA------------- 970

Qy    425 hrLeuArgAlaLeuTyrLysThrTyrIleAspCysVal---------HisValLysLysS 442
Db    971 --------TTAGTTGCTGATACTTACTTGGAAGCAATGTCTGTCACCCACTTCAAGAAAA 1022

Qy    442 erAspArgGlyArgLeuGlnThrGluAspProMetGluMetAspLysGluAsnAspMetT 462
Db   1023 AA---------------------------------------------------------- 1024

Qy    462 yrAlaGlyTyrHisGluSerAspThrSerGluAlaAlaAsnGluAlaLysIleGlnLysL 482
Db   1025 --------TATGAGGAGTATGATCTAAAAGGTGATGAACAAGAA------------CAAA 1064

Qy    482 euLysGluLeuSerLysLeuProGlyIleTyrAspArgLeuSerArgSerLeuAlaProS 502
Db   1065 TTGACCGATTGGCTGAGGACGGTGATATCTACAGTAAGCTGGCAAGGTCCTTGGCACCTG 1124

Qy    502 erIleTrpGluLeuGluAspIleLysLysGlyLeuLeuCysGlnLeuPheGlyGlyLysA 522
Db   1125 AAATATTTGGCCATGAAGATGTCAAAAAGGCACTGCTGTTACTACTTGTTGGTGCACCCC 1184

Qy    522 laLysLysIleProSerGlyAlaSerPheArgGlyAspIleAsnValLeuLeuValGlyA 542
Db   1185 ATCGGAAGCTCGCAGATGGCATGAAGATCAGAGGAGACCTGCATATATGCATGATGGGAG 1244

Qy    542 spProGlyThrSerLysSerGlnLeuLeuGlnTyrValHisLysIleAlaProArgGlyI 562
Db   1245 ATCCTGGTGTTGCAAAGAGTCAACTTCTGAAGCATATTATCAATGTTGCACCAAGAGGAG 1304

Qy    562 leTyrThrSerGlyArgGlySerSerAlaValGlyLeuThrAlaTyrValThrLysAspP 582
Db   1305 TGTACACCACTGGACGTGGGAGCAGTGGTGTTGGACTTACCGCTGCAGTCCAGAAAGATC 1364

Qy    582 roGluThrArgGluThrValLeuGluSerGlyAlaLeuValLeuSerAspArgGlyIleC 602
Db   1365 CAGTAACAAATGAGTTTGTCCTCGAAGGTGGAGCACTGGTACTGGCAGATATGGGCATTT 1424

Qy    602 ysCysIleAspGluPheAspLysMetSerAspAsnAlaArgSerMetLeuHisGluValM 622
Db   1425 GTGCTATAGATGAGTTTGACAAGATGGAAGAGTCAGACAGGACAGCGATTCATGAGGTAA 1484

Qy    622 etGluGlnGlnThrValSerValAlaLysGlyGlyIleIleAlaSerLeuAsnAlaArgT 642
Db   1485 TGGAGCAGCAAACAGTTAGCATTGCCAAGGCTGGCATCACCACCTCTCTTAATGCGAGAA 1544

Qy    642 hrSerValLeuAlaCysAlaAsnProSerGlySerArgTyrAsnAlaArgLeuSerValI 662
Db   1545 CTGCAATTCTGGCTGCTGCAAATCCAGCATGGGGAAGGTATGATATGAGGAGAACCCCAG 1604

Qy    662 leAspAsnIleGlnLeuProProThrLeuLeuSerArgPheAspLeuIleTyrLeuMetL 682
Db   1605 CAGAAAATATAAATCTACCTCCAGCTCTTCTGTCTCGTTTCGACCTCCTTTGGTTAATCC 1664

Qy    682 euAspLysProAspGluGlnAsnAspArgArgLeuAlaArgHisLeuValAlaLeuHisT 702
Db   1665 TGGATCGTGCAGACATGGAAACTGATCTTGAAATGGCAAGACACGTTGTTCATGTGCAT- 1723

Qy    702 yrGluAsnTyrGluValSerLysGln------AspAlaLeuAspLeuGlnThrLeuThrA 720
Db   1724 --CAAAATCTTGAATCACCAGCGCTGGGGTTCACACCACTTGAGCCATCTGTACTCAGAG 1781

Qy    720 laTyrIleThrTyrAlaArgGlnHisValHisProThrLeuSerAspGluAlaAlaGluA 740
Db   1782 CATACATATCTGCTGCTAGAAGA---GTCATTCCTTCTGTTCCTAGAGAGCTCGAGGAAT 1838
```

```
Qy        740 spLeuIleAsnGlyTyrValGluMetArgGln---------LysGlyAsnPheProGlyS 757
                 :::        |||       :::||||||         |||    |||    |||    |
Db       1839 ACATTGCAACTGCATATTCCAGCATCCGCCAAGAGGAGGCGAAGTCAAATGCACCAACCT 1898

Qy        757 erSerLysLysValIleThrAlaThrProArgGlnLeuGluSerMetIleArgIleSerG 777
                ||                          |||   |||   |||    |||:::::||||||||
Db       1899 CC---------------TACACAACTATCAGAACACTTTGAGCATACTCCGTATCTCTA 1943

Qy        777 luAlaLeuAlaArgMetArgPheSerGluValValGluLysValAspAlaAlaGluAlaV 797
                 |||||||||||:::||||||||||||   |||     :::     |||       ||||||:
Db       1944 TTGCCTTGGCAAGACTGAGGTTCTCAGAAACTGTGGCTCAGAGCGATGTCGACGAAGCAC 2003

Qy        797 alArgLeuLeuAspValAla-----------------------LeuGlnGlnSerAlaT 809
                ::|||||:::  ::::::                             |||:::|||
Db       2004 TGCGACTGATGCAAATGTCCAAGTACTCATTATACTCAGATGACCGCCAGCGGTCTGGCC 2063

Qy        809 hrAspHisAlaThrGlyThrIleAspMetAspLeuIleThrThrGlyValSerAlaSerG 829
                 |||          :::          :::                   ::::::|||
Db       2064 TTGATGCAATATCTGACATATATTCCATCCTGAGAGACGAAGCTGCTAGGACGAGCAGCA 2123

Qy        829 luArgIleArgArgAlaAsnLeuLeuAlaAlaLeuArgGluLeuIleAlaAspLysIleS 849
                    :::|||   |||:::   |||          :::
Db       2124 TGGATGTGAGATACGCTCACGCTCTTAACTTGATC------------------------ 2158

Qy        849 erProGlySerSerSerGlyLeuLysThrSerGlnLeuLeuGluAspIleArgSerGlnS 869
                          |||      |||          :::||||||   |||    :::          :
Db       2159 --------TCCAGAAAGGGATACAGCGAGGCTCAATTGAAGGAATGCTTGGAGGAATACG 2210

Qy        869 erSerValAspVal 873
                ::|||:::::|||
Db       2211 CTTCCCTGAATGTG 2224
```

RESULT 4
US-10-101-080-1
; Sequence 1, Application US/10101080
; Patent No. 6879916
; GENERAL INFORMATION:
; APPLICANT: Ramgopal Nadimpalli
; APPLICANT: Carl R. Simmons
; TITLE OF INVENTION: Maize Prolifera Gene and Uses Thereof
; FILE REFERENCE: 1138
; CURRENT APPLICATION NUMBER: US/10/101,080
; CURRENT FILING DATE: 2002-05-09
; PRIOR APPLICATION NUMBER: US/09/604,605
; PRIOR FILING DATE: 2000-06-27
; NUMBER OF SEQ ID NOS: 2
; SOFTWARE: FastSEQ for Windows Version 3.0
; SEQ ID NO 1
;   LENGTH: 2533
;   TYPE: DNA
;   ORGANISM: Zea mays
;   FEATURE:
;   NAME/KEY: CDS
;   LOCATION: (110)...(2269)
US-10-101-080-1

Alignment Scores:

Pred. No.:	9.8e-74	Length:	2533
Score:	993.00	Matches:	281
Percent Similarity:	49.88%	Conservative:	131
Best Local Similarity:	34.02%	Mismatches:	292
Query Match:	21.54%	Indels:	122
DB:	3	Gaps:	25

US-10-768-511-8 (1-901) x US-10-101-080-1 (1-2533)

```
Qy         84 ProTyrAspAlaGlyThrProGlyThrProGlyThrProValAlaThrProValTyrAla 103
              |||                   |||   |||          |||          |||:::::::::
Db          9 CCTTCTTCCCGCCATCCCCCGAACACTTCGCCGCCCCCCTCCCGCCTCCCGCTGTTCTCT 68

Qy        104 Thr-ProValGlyThrProMetGlyThrProSerPheHisArgGlyThrProGlnTyrLy 123
                  |||      ||||||          ||||||   ||||||          :::
Db         69 GCCTCCCACCCCCACGCCG---------CCGTCCCATCACCGCCGCTGGGATGAAGAACC 119

Qy        123 s------GlnArgSerGluLeuGlySerGlnGlyLysProLeuHisArgArgArgArgSe 141
                      :::|||   :::                |||||   |||:::           |||
Db        120 CTGACTTCGCCGCAGACAAAGCGCTCGCCAAGGACTTCCTGTCCAATTTCACCGG----- 174

Qy        141 rGlnSerArgGluProGlyHisArgSerProSerArgGluProSerAlaAspGlyArgPr 161
                                          |||||||||                        ||
Db        175 ---------------------------CCCTCACGGCGAGCCCAAGTACCTGAACATCC 206

Qy        161 oSerGluSerAlaGluProAspAspThrLeuGlyGlyGluTyrAlaTyrValTrpGlyTh 181
              |:::          |||    :::||||||  |||   :::   |||                |||
Db        207 TGCAAGATGTCGCGAACAGGAAGATACGCGCGGTCCAGATCGAGCT---------GGACG 257

Qy        181 rAsnValAsnIleProAspValLeuArgAlaIleArgArgPheLeuHisAsnTyrArgSe 201
                 :::   :::          |||   :::             |||                ::
Db        258 ACCTATTCCACTATAAGGATGTCGACGAGGAGTTCTTGCAGCGTGTCACCGAGAACACGC 317

Qy        201 rSerAlaHis---AspLeuAsnSerLysTyr-IleGlnIleIleGluGlu---------- 216
              ::::   |||         |||         :::     :::|||:::      |||
Db        318 GGCGCTACATCGGCATCTTCGCGGAGGCCATGGACGAGATCATGCCGGAGCCCACAGAGG 377

Qy        217 -----ThrValGluArgGluGluAspThrLeuAsnIleAspMetSerAspIleTyrAspH 235
                   ||||||:::  ::::::|||      |||         |||:::          |||
Db        378 CGTACACCGTCGACGAGGACCAGGACATCCTG---------ATGACGCAGCGTGTGGATG 428

Qy        235 isAspProAspLeuTyrAlaLysIleValArgTyrProLeuAspIleIleProLeuLeuA 255
                    |||      |||                                 ||||||
Db        429 AGGGGGCCGACGGCGGCGCGGAC-----------------GGCACGGACCCGCTA---- 466

Qy        255 spThrGluCysGlnGluValAlaThrSerLeuLeuProThrPheGluLysHisIleGluA 275
                          |||::::::         :::          ||||||    :::|||:::|
Db        467 -----------CAGAAAATGCCACCAGAGATCAAACGGTTCTTTGAGGTTTACATCAAGG 515

Qy        275 laArgProPheAsnLeuLysAlaSerValHisMetArgGluLeuAsnProSerAspIleA 295
              ||         |||:::          :::    :::|||::::::          |||:::|||
Db        516 CA------TTCTCAAAGGTGACTCCACTCACCATTCGGCAAGTGAAGGCATCCAACATTG 569

Qy        295 spLysLeuValSerValLysGlyMetValIleArgCysSerSerIleIleProGluIleL 315
                :::||||||    :::     |||:::|||     |||||||||     :::    |||    ::::
Db        570 GTCAGCTTGTGAAAATATCTGGGATTGTTACTCGCTGCTCGGATGTGAAGCCCTTGATGC 629

Qy        315 ysGlyAlaPhePheLysCysLeuValCysGlyHisSerProProLeuValThrValValL 335
              ::    |||    :::     |||        ||||||                  :::    |||
Db        630 AGGTTGCTGTTTATACATGTGAAGAATGTGGCTTCGAG------ATATACCAGGAAGTGA 683

Qy        335 ysGlyArgValGluGluPro---ThrArgCysGluLysProGluCys---------AlaA 351
                    ||||||          |||             |||                   |
Db        684 CTGCTAGAGTGTTTATGCCCCTCATTGAATGTCCATCTCAACGATGCAAGCTGAACAAAG 743

Qy        351 laArgAsnAlaMetSerLeuIleHisAsnArgCysThrPheAlaAsnLysGlnIleValA 371
              ||:::       :::    |||                         |||       |||     |||:
Db        744 CGAAGGGAAATCTAATCCTTCAACTGCGTGCGTCAAAGTTTTTGAAATTTCAGGAGGTAA 803

Qy        371 rgLeuGlnGluThrProAspAlaIleProGluGlyGluThrProHisThrValSerMetC 391
              ::|||||||||       :::    :::|||:::|||             |||       ::::::::
Db        804 AGCTTCAAGAGTTAGCAGAGCACGTACCAAAGGGCCACATTCCTCGTGCTCTGACTGTTC 863

Qy        391 ysLeuTyrAsnThrMetValAspAlaValLysProGlyAspArgIleGluValThrGlyV 411
                |||             :::          |||    |||||||||   :::|||::::::|||:
Db        864 ATCTAAGAGGAGAGCTGACTAGAAAGGTTGCACCTGGAGATGTGGTTGAGATGTCGGGTA 923

Qy        411 alPheLysAlaMet------------------AlaValArgValGlyProAsnGlnArgT 425
              ::|||       |||                   |||:::|||       |||
Db        924 TTTTTCTCCCTATGCCATACTACGGGTTTAGAGCGATGCGCGCAGGA------------- 970

Qy        425 hrLeuArgAlaLeuTyrLysThrTyrIleAspCysVal---------HisValLysLysS 442
                       :::         ||||||::::::    :::               |||     ||||||
Db        971 --------TTAGTTGCTGATACTTACTTGGAAGCAATGTCTGTCACCCACTTCAAGAAAA 1022

Qy        442 erAspArgGlyArgLeuGlnThrGluAspProMetGluMetAspLysGluAsnAspMetT 462

Db       1023 AA---------------------------------------------------------- 1024
```

```
Qy    462 yrAlaGlyTyrHisGluSerAspThrSerGluAlaAlaAsnGluAlaLysIleGlnLysL 482
Db   1025 --------TATGAGGAGTATGATCTAAAAGGTGATGAACAAGAA------------CAAA 1064

Qy    482 euLysGluLeuSerLysLeuProGlyIleTyrAspArgLeuSerArgSerLeuAlaProS 502
Db   1065 TTGACCGATTGGCTGAGGACGGTGATATCTACAGTAAGCTGGCAAGGTCCTTGGCACCTG 1124

Qy    502 erIleTrpGluLeuGluAspIleLysLysGlyLeuLeuCysGlnLeuPheGlyGlyLysA 522
Db   1125 AAATATTTGGCCATGAAGATGTCAAAAAGGCACTGCTGTTACTACTTGTTGGTGCACCCC 1184

Qy    522 laLysLysIleProSerGlyAlaSerPheArgGlyAspIleAsnValLeuLeuValGlyA 542
Db   1185 ATCGGAAGCTCGCAGATGGCATGAAGATCAGAGGAGACCTGCATATATGCATGATGGGAG 1244

Qy    542 spProGlyThrSerLysSerGlnLeuLeuGlnTyrValHisLysIleAlaProArgGlyI 562
Db   1245 ATCCTGGTGTTGCAAAGAGTCAACTTCTGAAGCATATTATCAATGTTGCACCAAGAGGAG 1304

Qy    562 leTyrThrSerGlyArgGlySerSerAlaValGlyLeuThrAlaTyrValThrLysAspP 582
Db   1305 TGTACACCACTGGACGTGGGAGCAGTGGTGTTGGACTTACCGCTGCAGTCCAGAAAGATC 1364

Qy    582 roGluThrArgGluThrValLeuGluSerGlyAlaLeuValLeuSerAspArgGlyIleC 602
Db   1365 CAGTAACAAATGAGTTTGTCCTCGAAGGTGGAGCACTGGTACTGGCAGATATGGGCATTT 1424

Qy    602 ysCysIleAspGluPheAspLysMetSerAspAsnAlaArgSerMetLeuHisGluValM 622
Db   1425 GTGCTATAGATGAGTTTGACAAGATGGAAGAGTCAGACAGGACAGCGATTCATGAGGTAA 1484

Qy    622 etGluGlnGlnThrValSerValAlaLysGlyGlyIleIleAlaSerLeuAsnAlaArgT 642
Db   1485 TGGAGCAGCAAACAGTTAGCATTGCCAAGGCTGGCATCACCACCTCTCTTAATGCGAGAA 1544

Qy    642 hrSerValLeuAlaCysAlaAsnProSerGlySerArgTyrAsnAlaArgLeuSerValI 662
Db   1545 CTGCAATTCTGGCTGCTGCAAATCCAGCATGGGGAAGGTATGATATGAGGAGAACCCCAG 1604

Qy    662 leAspAsnIleGlnLeuProProThrLeuLeuSerArgPheAspLeuIleTyrLeuMetL 682
Db   1605 CAGAAAATATAAATCTACCTCCAGCTCTTCTGTCTCGTTTCGACCTCCTTTGGTTAATCC 1664

Qy    682 euAspLysProAspGluGlnAsnAspArgArgLeuAlaArgHisLeuValAlaLeuHisT 702
Db   1665 TGGATCGTGCAGACATGGAAACTGATCTTGAAATGGCAAGACACGTTGTTCATGTGCAT- 1723

Qy    702 yrGluAsnTyrGluValSerLysGln------AspAlaLeuAspLeuGlnThrLeuThrA 720
Db   1724 --CAAAATCTTGAATCACCAGCGCTGGGGTTCACACCACTTGAGCCATCTGTACTCAGAG 1781

Qy    720 laTyrIleThrTyrAlaArgGlnHisValHisProThrLeuSerAspGluAlaAlaGluA 740
Db   1782 CATACATATCTGCTGCTAGAAGA---GTCATTCCTTCTGTTCCTAGAGAGCTCGAGGAAT 1838

Qy    740 spLeuIleAsnGlyTyrValGluMetArgGln---------LysGlyAsnPheProGlyS 757
Db   1839 ACATTGCAACTGCATATTCCAGCATCCGCCAAGAGGAGGCGAAGTCAAATGCACCAACCT 1898

Qy    757 erSerLysLysValIleThrAlaThrProArgGlnLeuGluSerMetIleArgIleSerG 777
Db   1899 CC---------------TACACAACTATCAGAACACTTTTGAGCATACTCCGTATCTCTA 1943

Qy    777 luAlaLeuAlaArgMetArgPheSerGluValValGluLysValAspAlaAlaGluAlaV 797
Db   1944 TTGCCTTGGCAAGACTGAGGTTCTCAGAAACTGTGGCTCAGAGCGATGTCGACGAAGCAC 2003

Qy    797 alArgLeuLeuAspValAla-----------------------LeuGlnGlnSerAlaT 809
Db   2004 TGCGACTGATGCAAATGTCCAAGTACTCATTATACTCAGATGACCGCCAGCGGTCTGGCC 2063

Qy    809 hrAspHisAlaThrGlyThrIleAspMetAspLeuIleThrThrGlyValSerAlaSerG 829
Db   2064 TTGATGCAATATCTGACATATATTCCATCCTGAGAGACGAAGCTGCTAGGACGAGCAGCA 2123

Qy    829 luArgIleArgArgAlaAsnLeuLeuAlaAlaLeuArgGluLeuIleAlaAspLysIleS 849
Db   2124 TGGATGTGAGATACGCTCACGCTCTTAACTTGATC------------------------- 2158

Qy    849 erProGlySerSerSerGlyLeuLysThrSerGlnLeuLeuGluAspIleArgSerGlnS 869
Db   2159 --------TCCAGAAAGGGATACAGCGAGGCTCAATTGAAGGAATGCTTGGAGGAATACG 2210

Qy    869 erSerValAspVal 873
Db   2211 CTTCCCTGAATGTG 2224
```

RESULT 5
US-09-949-016-2621
; Sequence 2621, Application US/09949016
; Patent No. 6812339
; GENERAL INFORMATION:
; APPLICANT: VENTER, J. Craig et al.
; TITLE OF INVENTION: POLYMORPHISMS IN KNOWN GENES ASSOCIATED
; TITLE OF INVENTION: WITH HUMAN DISEASE, METHODS OF DETECTION AND USES THEREOF
; FILE REFERENCE: CL001307
; CURRENT APPLICATION NUMBER: US/09/949,016
; CURRENT FILING DATE: 2000-04-14
; PRIOR APPLICATION NUMBER: 60/241,755
; PRIOR FILING DATE: 2000-10-20
; PRIOR APPLICATION NUMBER: 60/237,768
; PRIOR FILING DATE: 2000-10-03
; PRIOR APPLICATION NUMBER: 60/231,498
; PRIOR FILING DATE: 2000-09-08
; NUMBER OF SEQ ID NOS: 207012
; SOFTWARE: FastSEQ for Windows Version 4.0
; SEQ ID NO 2621
;   LENGTH: 3000
;   TYPE: DNA
;   ORGANISM: Human
US-09-949-016-2621

Alignment Scores:

Pred. No.:	9.97e-73	Length:	3000
Score:	982.50	Matches:	299
Percent Similarity:	45.41%	Conservative:	156
Best Local Similarity:	29.84%	Mismatches:	304
Query Match:	21.32%	Indels:	244
DB:	3	Gaps:	36

US-10-768-511-8 (1-901) x US-09-949-016-2621 (1-3000)

```
Qy     67 ArgArgArgArgSerArgSerArgSerLeuGlyAsnSerValTyrSerSerProTyrAsp 86
Db      7 CGGAACAAAAGAACGCGTGTGCGCTGGCCCTTTAAGAGCGA-TTCTCCTCCGCCCGCGCC 65

Qy     87 AlaGlyThrProGlyThrProGlyThrProValAlaThrProValTyrAlaThrProVal 106
Db     66 AGCTCGGACCGCGGGAAACCCGGCGCCTGCACTACCCCGCCCGGAGATTCCCTTCCGACG 125

Qy    107 GlyThrProMetGlyThrProSerPheHisArgGlyThr-------------------- 119
Db    126 CCCGCACCGCCTCCCCGTCACTCATTCTAGGCCCGCACGGTGATTGGCTTGCGGCTAGCG 185

Qy    120 ------------------------------------ProGlnTyrLysGlnArgSerGlu 127
Db    186 GGAGGTGAAGAAGGCCGCCTTGTCCGATTGGCCCGCACGCAGTGGCGC---CGGTCACGT 242

Qy    128 LeuGly------------------SerGlnGlyLysProLeuHisArgArgArgArgSer 141
Db    243 GGGGGGCGACGTTTCGCGCCAATTTCGGTTGGCCGGCCACAGTCCACCGCGCGGAGATTC 302

Qy    142 GlnSerArgGluProGlyHisArgSerProSerArgGluProSerAlaAspGlyArgPro 161
```

```
Db        303 ---TCAGCTTCCCCAGGAGCAAGA---CCTCTGAGCCCGCCAAGCGCGGCCGCACGGCCC 356
Qy        162 SerGluSerAlaGluProAspAspThrLeuGlyGlyGluTyrAlaTyrValTrpGlyThr 181
Db        357 TCGGCAGCGATGGCACTGAAGGAC---------------TACGCGCTAGAGAAGGGTACG 401
Qy        182 AsnValAsnIleProAsp---ValLeuArgAla-------------------------- 191
Db        402 GGTCTACGAGCCCGGGAGGGCGTCCTGCGCGGGGAGCCTCCCGGGGAACACCTGTTGTTC 461
Qy        191 ------------------------------------------------------------ 191
Db        462 ATAGGTCGTGTTTTAAGCGAAAATGGGTATCTAAGATTTGTCTTAAAGTTTGTGTCTTAC 521
Qy        192 ---------------------------IleArgArgPheLeuHisAsnTyrArgSerSer 202
Db        522 GGTTTTACATTTCATGTTACAGAAAAGGTTAAGAAGTTCTTACAAGAGTTCTACCAGGAT 581
Qy        203 AlaHisAspLeuAsnSerLysTyrIleGlnIleIleGluGluThrVal---------Glu 219
Db        582 ---GATGAACTCGGGAAGAAGCAGTTCAAGTATGGGAACCAGTTGGTTCGGCTGGCTCAT 638
Qy        220 ArgGluGluAspThrLeuAsnIleAspMetSerAspIleTyrAspHisAspProAspLeu 239
Db        639 CGGGAACAGGTGGCTCTGTATGTGGACCTGGACGACGTAGCCGAGGATGACCCCGAGTTG 698
Qy        240 ------------------------------TyrAlaLysIleValArgTyrProLeu--- 248
Db        699 GTGGACTCAATTTGTGAGAATGCCAGGCGCTACGCGAAGCTCTTTGCTGATGCCGTACAA 758
Qy        249 AspIleIlePro--------------------------------LeuLeuAspThrGlu 257
Db        759 GAGCTGCTGCCTCAGTACAAGGAGAGGGAAGTGGTAAATAAAGATGTCCTGGACGTTTAC 818
Qy        258 CysGln------------------------------------------------------ 259
Db        819 ATTGAGCATCGGCTAATGATGGAGCAGCGGAGTCGGGACCCTGGGATGGTCCGAAGCCCC 878
Qy        260 ------GluValAlaThrSerLeuLeuProThrPheGluLysHisIleGluAlaArgPro 277
Db        879 CAGAACCAGTACCCTGCTGAACTCATGCGCAGATTTGAGCTGTATTTTCAAGGC---CCT 935
Qy        278 PheAsnLeuLysAlaSerValHisMetArgGluLeuAsnProSerAspIleAspLysLeu 297
Db        936 AGCAGCAACAAGCCTCGTGTG---ATCCGGGAAGTGCGGCTGACTCTGTGGGGAAGTTG 992
Qy        298 ValSerValLysGlyMetValIleArgCysSerSerIleIleProGluIleLysGlyAla 317
Db        993 GTAACTGTGCGTGGAATCGTCACTCGTGTCTCTGAAGTCAAACCCAAGATGGTGGTGGCC 1052
Qy        318 PhePheLysCysLeuValCysGly---------HisSerProProLeuValThrValVal 334
Db       1053 ACTTACACTTGTGACCAGTGTGGGGCAGAGACCTACCAGCCG------------------ 1094
Qy        335 LysGlyArgValGluGluProThr---------------ArgCysGluLysProGluCys 349
Db       1095 ---------ATCCAGTCTCCCACTTTCATGCCTCTGATCATGTGCCCAAGCCAGGAGTGC 1145
Qy        350 AlaAlaArgAsnAla---------MetSerLeuIleHisAsnArgCysThrPheAlaAsn 366
Db       1146 CAAACCAACCGCTCAGGAGGGCGGCTGTATCTGCAGACACGGGGCTCCAGATTCATCAAA 1205
Qy        367 LysGlnIleValArgLeuGlnGluThrProAspAlaIleProGluGlyGluThrProHis 386
Db       1206 TTCCAGGAGATGAAGATGCAAGAACATAGTGATCAGGTGCCTGTGGGAAATATCCCTCGT 1265
Qy        387 ThrValSerMetCysLeuTyrAsnThrMetValAspAlaValLysProGlyAspArgIle 406
Db       1266 AGTATCACGGTGCTGGTAGAAGGAGAGAACACAAGGATTGCCCAGCCTGGAGACCACGTC 1325
Qy        407 GluValThrGlyValPheLysAlaMetAlaValArgValGly------------------ 420
Db       1326 AGCGTCACTGGTATTTTCTTGCCAATC---CTGCGCACTGGGTTCCGACAGGTGGTACAG 1382
Qy        421 ------------------ProAsnGln---ArgThr------------------------ 425
Db       1383 AGAAAAGAGCAGCAAAACCCCAAACAAAACAGAACCCAGGGATTTGTCAAGAAATTGTTA 1442
Qy        426 ---------------------LeuArgAlaLeuTyr------------------------ 430
Db       1443 GGGAATCCAGTGATGGGCAAGTTGAAGGGATTACATCTGTATCCTGGCATTTTCCTGCCC 1502
Qy        431 ---------------LysThrTyrIleAspCysValHisValLysLysSerAspArgGly 445
Db       1503 AAGGGTTTACTCTCAGAAACCTACCTGGAAGCCCATCGGATTGTGAAG------------ 1550
Qy        446 ArgLeuGlnThrGluAspProMetGluMetAspLysGluAsnAspMetTyrAlaGlyTyr 465
Db       1551 ---------------------------ATGAACAAGAGTGAGGATGATGAGTCTGGG--- 1580
Qy        466 HisGluSerAspThrSerGluAlaAlaAsnGluAlaLysIleGlnLysLeuLysGluLeu 485
Db       1581 ------------------------GCTGGAGAGCTCACCAGGGAGGAGCTGAGGCAAATT 1616
Qy        486 SerLysLeuProGlyIleTyrAspArgLeuSerArgSerLeuAlaProSerIleTrpGlu 505
Db       1617 GCAGAG---GAGGATTTCTACGAAAAGCTGGCAGCTTCAATCGCCCCAGAAATATACGGG 1673
Qy        506 LeuGluAspIleLysLysGlyLeuLeuCysGlnLeuPheGlyGlyLysAlaLysLysIle 525
Db       1674 CATGAAGATGTGAAGAAGGCACTGCTGCTCCTGCTAGTCGGGGGT---GTGGACCAGTCT 1730
Qy        526 ProSerGlyAlaSerPheArgGlyAspIleAsnValLeuLeuValGlyAspProGlyThr 545
Db       1731 CCTCGAGGCATGAAAATCCGGGGCAACATCAACATCTGTCTGATGGGGGATCCTGGTGTG 1790
Qy        546 SerLysSerGlnLeuLeuGlnTyrValHisLysIleAlaProArgGlyIleTyrThrSer 565
Db       1791 GCCAAGTCTCAGCTCCTGTCATACATTGATCGACTGGCGCCTCGCAGCCAGTACACAACA 1850
Qy        566 GlyArgGlySerSerAlaValGlyLeuThrAlaTyrValThrLysAspProGluThrArg 585
Db       1851 GGCCGGGGCTCCTCAGGAGTGGGGCTTACGGCAGCTGTGCTGAGAGACTCCGTGAGTGGA 1910
Qy        586 GluThrValLeuGluSerGlyAlaLeuValLeuSerAspArgGlyIleCysCysIleAsp 605
Db       1911 GAACTGACCTTAGAGGGTGGGGCCCTGGTGCTGGCTGACCAGGGTGTGTGCTGCATTGAT 1970
Qy        606 GluPheAspLysMetSerAspAsnAlaArgSerMetLeuHisGluValMetGluGlnGln 625
Db       1971 GAGTTCGACAAGATGGCTGAGGCCGACCGCACAGCCATCCACGAGGTCATGGAGCAGCAG 2030
Qy        626 ThrValSerValAlaLysGlyGlyIleIleAlaSerLeuAsnAlaArgThrSerValLeu 645
Db       2031 ACCATCTCCATTGCCAAGGCCGGCATTCTCACCACACTCAATGCCCGCTGCTCCATCCTG 2090
Qy        646 AlaCysAlaAsnProSerGlySerArgTyrAsnAlaArgLeuSerValIleAspAsnIle 665
Db       2091 GCTGCCGCCAACCCTGCCTACGGGCGCTACAACCCTCGCCGCAGCCTGGAGCAGAACATA 2150
Qy        666 GlnLeuProProThrLeuLeuSerArgPheAspLeuIleTyrLeuMetLeuAspLysPro 685
Db       2151 CAGCTACCTGCTGCACTGCTCTCCCGGTTTGACCTCCTCTGGCTGATTCAGGACCGGCCC 2210
Qy        686 AspGluGlnAsnAspArgArgLeuAlaArgHisLeuValAlaLeuHisTyrGluAsnTyr 705
Db       2211 GACCGAGACAATGACCTACGGTTGGCCCAGCACATCACCTATGTGCACCAGCACAGCCGG 2270
Qy        706 GluValSerLysGln---AspAlaLeuAspLeuGlnThrLeuThrAlaTyrIleThrTyr 724
Db       2271 CAGCCCCCCTCCCAGTTTGAACCTCTGGACATGAAGCTCATGAGGCGTTACATAGCCATG 2330
Qy        725 AlaArgGlnHisValHisProThrLeuSerAspGluAlaAlaGluAspLeuIleAsnGly 744
Db       2331 TGCCGCGAGAAG---CAGCCCATGGTGCCAGAGTCTCTGGCTGACTACATCACAGCAGCA 2387
```

```
Qy        745 TyrValGluMetArgGlnLysGlyAsnPheProGlySerSerLysLysValIleThrAla 764
Db       2388 TACGTGGAGATGAGGCGAGAGGCT---------TGGGCTAGTAAGGATGCCACCTATACT 2438

Qy        765 ThrProArgGlnLeuGluSerMetIleArgIleSerGluAlaLeuAlaArgMetArgPhe 784
Db       2439 TCTGCCCGGACCCTGCTGGCTATCCTGCGCCTTTCCACTGCTCTGGCACGTCTGAGAATG 2498

Qy        785 SerGluValValGluLysValAspAlaAlaGluAlaValArgLeuLeuAspValAlaLeu 804
Db       2499 GTGGATGTGGTGGAGAAAGAAGATGTGAATGAAGCCATCAGGCTAATGGAGATGTCAAAG 2558

Qy        805 GlnGlnSerAlaThrAspHisAlaThrGlyThrIleAspMetAspLeuIleThrThrGly 824
Db       2559 GACTCTCTTCTAGGAGAC----------------------------------AAGGGG 2582

Qy        825 ValSerAlaSerGluArgIleArgArgAlaAsnLeuLeuAlaAlaLeuArgGluLeuIle 844
Db       2583 CAGACAGCTAGGACTCAGAGACCAGCAGATGTGATATTTGCCACCGTCCGTGAACTG--- 2639

Qy        845 AlaAspLysIleSerProGlySerSer-SerGlyLeuLysThrSerGlnLeuLeuGluAs 864
Db       2640 ---------GTCTCAGGGGGCCGAAGTGTCCGGT-----------------------TC 2666

Qy        864 pIleArgSerGlnSerSerValAspVal--------------------SerLeuGlnAs 877
Db       2667 TCTGAGGCAGAGCAGCGCTGTGTATCTCGTGGCTTCACACCCGCCCAGTTCCAGGCGGCT 2726

Qy        877 pIleLysAsnAlaLeuGlySerLeuGlnGlyGluGlyPheLeuThrValHisGlyAspIl 897
Db       2727 CTGGATGAATATGAGGAGCTCAATGTCTGGCAGGTCAATGCTTCCCGGACACGGATCACT 2786

Qy        897 eVal 898
Db       2787 TTTG 2790
```

RESULT 6
US-09-964-899-48
; Sequence 48, Application US/09964899
; Patent No. 6900367
; GENERAL INFORMATION:
; APPLICANT: Cohen, Dalia et al.
; TITLE OF INVENTION: Identification of Genes Involved in
; TITLE OF INVENTION: Alzheimer's Disease Using Drosophila Melanogaster
; FILE REFERENCE: 4-31612 A
; CURRENT APPLICATION NUMBER: US/09/964,899
; CURRENT FILING DATE: 2001-09-27
; PRIOR APPLICATION NUMBER: 60/236,893
; PRIOR FILING DATE: 2000-09-29
; PRIOR APPLICATION NUMBER: 60/298,309
; PRIOR FILING DATE: 2001-06-14
; NUMBER OF SEQ ID NOS: 53
; SOFTWARE: FastSEQ for Windows Version 4.0
; SEQ ID NO 48
; LENGTH: 3378
; TYPE: DNA
; ORGANISM: Homo Sapien
US-09-964-899-48

Alignment Scores:

Pred. No.:	1.34e-69	Length:	3378
Score:	946.50	Matches:	262
Percent Similarity:	46.45%	Conservative:	157
Best Local Similarity:	29.05%	Mismatches:	344
Query Match:	20.54%	Indels:	139
DB:	3	Gaps:	23

US-10-768-511-8 (1-901) x US-09-964-899-48 (1-3378)

```
Qy         29 GlnValThrSerProSerPheAspAsnAlaAlaSerProValAlaGlyArgArgAlaVal 48
Db          4 CAGGAATCATCGGAATCCTTCACCATGGCATCCAGCCCG---GCCCAGCGTCGGCGAGGC 60

Qy         49 ArgGlnThrProThrSerAlaValArgArgArgGlyArgGluThrAspSerAlaArgArg 68
Db         61 AATGATCCTCTCACCTCCAGCCCTGGCCGAAGCTCCCGGCGTACTGATGCCCTC------ 114

Qy         69 ArgArgSerArgSerArgSerLeuGlyAsnSerValTyrSerSerProTyrAspAlaGly 88
Db        115 ACCTCCAGCCCTGGCCGTGACCTT------------------CCACCATTTGAGGATGAG 156

Qy         89 ThrProGlyThrProGlyThrProValAlaThrProValTyrAlaThrProValGlyThr 108
Db        157 TCCGAGGGGCTCCTAGGCACA------GAGGGGCCCCTGGAGGAAGAAGAGGATGGAGAG 210

Qy        109 ProMet---------GlyThrProSerPheHisArgGlyThrProGlnTyrLysGlnArg 125
Db        211 GAGCTCATTGGAGATGGCATGGAAAGGGACTACCGCGCCATCCCAGAGCTGGACGCCTAT 270

Qy        126 SerGluLeuGly---------------------------------------SerGlnGly 132
Db        271 GAGGCCGAGGGACTGGCTCTGGATGATGAGGACGTAGAGGAGCTGACGGCCAGTCAGAGG 330

Qy        133 LysProLeuHisArgArgArgArgSerGlnSerArgGluProGly--------------- 147
Db        331 GAGGCAGCAGAGCGGGCCATGCGGCAGCGTGACCGGGAGGCTGGCCGGGGCCTGGGCCGC 390

Qy        148 ---------------------------------------HisArgSerProSerArgGlu 154
Db        391 ATGCGCCGTGGGCTCCTGTATGACAGCGATGAGGAGGACGAGGAGCGCCCTGCCCGCAAG 450

Qy        155 ProSerAlaAspGlyArgProSerGluSerAlaGluProAsp------------------ 168
Db        451 CGCCGCCAGGTGGAGCGGGCCACGGAGGACGGCGAGGAGGACGAGGAGATGATTGAGAGC 510

Qy        169 ------------AspThrLeuGlyGlyGluTyrAlaTyrValTrp------------Gly 180
Db        511 ATCGAGAACCTGGAGGATCTCAAAGGCCACTCTGTGCGCGAGTGGGTGAGCATGGCGGGC 570

Qy        181 ThrAsnValAsnIleProAspValLeuArgAlaIleArgArgPheLeuHisAsnTyr--- 199
Db        571 CCCCGGCTG---------GAGATCCACCACCGCTTCAAGAACTTCCTGCGCACTCACGTC 621

Qy        200 ArgSerSerAlaHisAspLeuAsnSerLysTyrIleGlnIleIleGluGluThrValGlu 219
Db        622 GACAGCCACGGCCACAACGTC---------TTCAAGGAGCGCATCAGCGACATGTGCAAA 672

Qy        220 ArgGluGluAspThrLeuAsnIleAspMetSerAspIleTyrAspHisAspProAspLeu 239
Db        673 GAGAACCGTGAGAGCCTGGTGGTGAACTATGAGGACTTGGCAGCCAGGGAGCACGTGCTG 732

Qy        240 TyrAlaLysIleValArgTyrProLeuAspIleIleProLeuLeuAspThrGluCysGln 259
Db        733 GCCTACTTCCTGCCTGAGGCACCGGCGGAGCTGCTGCAGATCTTTGATGAGGCTGCCCTG 792

Qy        260 GluValAlaThrSerLeuLeuProThrPheGluLys---------HisIleGluAlaArg 276
Db        793 GAGGTGGTACTGGCCATGTACCCCAAGTACGACCGCATCACCAACCACATCCATGTCCGC 852

Qy        277 ProPheAsnLeuLysAlaSerValHisMetArgGluLeuAsnProSerAspIleAspLys 296
Db        853 ATCTCCCACCTGCCTCTGGTGGAGGAGCTGCGCTCGCTGAGGCAGCTGCATCTGAACCAG 912

Qy        297 LeuValSerValLysGlyMetValIleArgCysSerSerIleIleProGluIleLysGly 316
Db        913 CTGATCCGCACCAGTGGGGTGGTGACCAGCTGCACTGGCGTCCTGCCCCAGCTCAGCATG 972

Qy        317 AlaPhePheLysCysLeuValCysGlyHisSerProProLeuValThrValValLysGly 336
Db        973 GTCAAGTACAACTGCAACAAGTGCAATTTCGTCCTGGGTCCTTTCTGCCAGTCCCAGAAC 1032

Qy        337 ArgValGluGluProThrArgCysGluLysProGluCysAlaAlaArgAsnAlaMetSer 356
Db       1033 CAGGAGGTGAAACCAGGCTCCTGT------CCTGAGTGCCAGTCGGCCGGCCCCTTTGAG 1086

Qy        357 LeuIleHisAsnArgCysThrPheAlaAsnLysGlnIleValArgLeuGlnGluThrPro 376
```

```
Db       1087 GTCAACATGGAGGAGACCATCTATCAGAACTACCAGCGTATCCGAATCCAGGAGAGTCCA 1146

Qy        377 AspAlaIleProGluGlyGluThrProHisThrValSerMetCysLeuTyrAsnThrMet 396
Db       1147 GGCAAAGTGGCGGCTGGCCGGCTGCCCCGCTCCAAGGACGCCATTCTCCTCGCAGATCTG 1206

Qy        397 ValAspAlaValLysProGlyAspArgIleGluValThrGlyValPheLysAlaMetAla 416
Db       1207 GTGGACAGCTGCAAGCCAGGAGACGAGATAGAGCTGACTGGCATCTATCACAAC------ 1260

Qy        417 ValArgValGlyProAsnGlnArgThrLeuArgAla------LeuTyrLysThrTyrIle 434
Db       1261 ---AACTATGATGGCTCCCTCAACACTGCCAATGGCTTCCCTGTCTTTGCCACTGTCATC 1317

Qy        435 AspCysValHisValLysLysSerAspArg-----------GlyArgLeuGlnThrGlu 450
Db       1318 CTAGCCAACCACGTGGCCAAGAAGGACAACAAGGTTGCTGTAGGGGAACTGACCGATGAA 1377

Qy        451 AspProMetGluMetAspLysGluAsnAspMetTyrAlaGlyTyrHisGluSerAspThr 470
Db       1378 GAT--------------------------------------------------------- 1380

Qy        471 SerGluAlaAlaAsnGluAlaLysIleGlnLysLeuLysGluLeuSerLysLeuProGly 490
Db       1381 ------------------------GTGAAGATGATCACTAGCCTCTCCAAGGATCAGCAG 1416

Qy        491 IleTyrAspArgLeuSerArgSerLeuAlaProSerIleTrpGluLeuGluAspIleLys 510
Db       1417 ATCGGAGAGAAGATCTTTGCCAGCATTGCTCCTTCCATCTATGGTCATGAAGACATCAAG 1476

Qy        511 LysGlyLeuLeuCysGlnLeuPheGlyGlyLysAlaLysLysIleProSerGlyAlaSer 530
Db       1477 AGAGGCCTGGCTCTGGCCCTGTTCGGAGGGGAGCCCAAAAACCCAGGTGGCAAGCACAAG 1536

Qy        531 PheArgGlyAspIleAsnValLeuLeuValGlyAspProGlyThrSerLysSerGlnLeu 550
Db       1537 GTACGTGGTGATATCAACGTGCTCTTGTGCGGAGACCCTGGCACAGCGAAGTCGCAGTTT 1596

Qy        551 LeuGlnTyrValHisLysIleAlaProArgGlyIleTyrThrSerGlyArgGlySerSer 570
Db       1597 CTCAAGTATATTGAGAAAGTGTCCAGCCGAGCCATCTTCACCACTGGCCAGGGGGCGTCG 1656

Qy        571 AlaValGlyLeuThrAlaTyrValThrLysAspProGluThrArgGluThrValLeuGlu 590
Db       1657 GCTGTGGGCCTCACGGCGTATGTCCAGCGGCACCCTGTCAGCAGGGAGTGGACCTTGGAG 1716

Qy        591 SerGlyAlaLeuValLeuSerAspArgGlyIleCysCysIleAspGluPheAspLysMet 610
Db       1717 GCTGGGGCCCTGGTTCTGGCTGACCGAGGAGTGTGTCTCATTGATGAATTTGACAAGATG 1776

Qy        611 SerAspAsnAlaArgSerMetLeuHisGluValMetGluGlnGlnThrValSerValAla 630
Db       1777 AATGACCAGGACAGAACCAGCATCCATGAGGCCATGGAGCAACAGAGCATCTCCATCTCG 1836

Qy        631 LysGlyGlyIleIleAlaSerLeuAsnAlaArgThrSerValLeuAlaCysAlaAsnPro 650
Db       1837 AAGGCTGGCATCGTCACCTCCCTGCAGGCTCGCTGCACGGTCATTGCTGCCGCCAACCCC 1896

Qy        651 SerGlySerArgTyrAsnAlaArgLeuSerValIleAspAsnIleGlnLeuProProThr 670
Db       1897 ATAGGAGGGCGCTACGACCCCTCGCTGACTTTCTCTGAGAACGTGGACCTCACAGAGCCC 1956

Qy        671 LeuLeuSerArgPheAspLeuIleTyrLeuMetLeuAspLysProAspGluGlnAsnAsp 690
Db       1957 ATCATCTCACGCTTTGACATCCTGTGTGTGGTGAGGGACACCGTGGACCCAGTCCAGGAC 2016

Qy        691 ArgArgLeuAlaArgHisLeuValAlaLeuHisTyrGluAsnTyrGluValSerLysGln 710
Db       2017 GAGATGCTGGCCCGCTTCGTGGTGGGCAGCCACGTCAGACACCACCCCAGCAACAAGGAG 2076

Qy        710 ------------------------------------------------------------ 710
Db       2077 GAGGAGGGGCTGGCCAATGGCAGCGCTGCTGAGCCCGCCATGCCCAACACGTATGGCGTG 2136

Qy        711 AspAlaLeuAspLeuGlnThrLeuThrAlaTyrIleThrTyrAlaArgGlnHisValHis 730
Db       2137 GAGCCCCTGCCCCAGGAGGTCCTGAAGAAGTACATCATCTACGCCAAGGAGAGGGTCCAC 2196

Qy        731 ProThrLeuSerAspGluAlaAlaGluAspLeuIleAsnGlyTyrValGluMetArgGln 750
Db       2197 CCGAAGCTCAACCAGATGGACCAGGACAAGGTGGCCAAGATGTACAGTGACCTGAGGAAA 2256

Qy        751 LysGlyAsnPheProGlySerSerLysLysValIleThrAlaThrProArgGlnLeuGlu 770
Db       2257 GAATCTATGGCGACAGGCAGC------------ATCCCCATTACGGTGCGGCACATCGAG 2304

Qy        771 SerMetIleArgIleSerGluAlaLeuAlaArgMetArgPheSerGluValValGluLys 790
Db       2305 TCCATGATCCGCATGGCGGAGGCCCACGCGCGCATCCATCTGCGGGACTATGTGATCGAA 2364

Qy        791 ValAspAlaAlaGluAlaValArgLeuLeuAspValAlaLeuGlnGlnSerAlaThrAsp 810
Db       2365 GACGACGTCAACATGGCCATCCGCGTG------------ATGCTGGAGAGCTTCATAGAC 2412

Qy        811 HisAlaThrGlyThrIleAspMetAspLeuIleThrThrGlyValSerAlaSerGluArg 830
Db       2413 ACACAGAAGTTCAGCGTCATGCGCAGCATGCGCAAGACT---TTTGCCCGCTACCTTTCA 2469

Qy        831 IleArgArgAlaAsn---------LeuLeuAlaAlaLeuArgGluLeuIleAlaAspLys 847
Db       2470 TTCCGGCGTGACAACAATGAGCTGTTGCTCTTCATACTGAAGCAGTTAGTGGCAGAGCAG 2529

Qy        848 IleSer 849
Db       2530 GTGACA 2535
```

RESULT 7
US-09-949-016-5021
; Sequence 5021, Application US/09949016
; Patent No. 6812339
; GENERAL INFORMATION:
; APPLICANT: VENTER, J. Craig et al.
; TITLE OF INVENTION: POLYMORPHISMS IN KNOWN GENES ASSOCIATED
; TITLE OF INVENTION: WITH HUMAN DISEASE, METHODS OF DETECTION AND USES THEREOF
; FILE REFERENCE: CL001307
; CURRENT APPLICATION NUMBER: US/09/949,016
; CURRENT FILING DATE: 2000-04-14
; PRIOR APPLICATION NUMBER: 60/241,755
; PRIOR FILING DATE: 2000-10-20
; PRIOR APPLICATION NUMBER: 60/237,768
; PRIOR FILING DATE: 2000-10-03
; PRIOR APPLICATION NUMBER: 60/231,498
; PRIOR FILING DATE: 2000-09-08
; NUMBER OF SEQ ID NOS: 207012
; SOFTWARE: FastSEQ for Windows Version 4.0
; SEQ ID NO 5021
; LENGTH: 2533
; TYPE: DNA
; ORGANISM: Human
US-09-949-016-5021

Alignment Scores:

Pred. No.:	3.61e-69	Length:	2533
Score:	939.00	Matches:	232
Percent Similarity:	51.73%	Conservative:	156
Best Local Similarity:	30.93%	Mismatches:	290
Query Match:	20.37%	Indels:	72
DB:	3	Gaps:	18

US-10-768-511-8 (1-901) x US-09-949-016-5021 (1-2533)

```
Qy        169 AspThrLeuGlyGlyGluTyrAlaTyrValTrpGlyThrAsnValAsnIleProAspVal 188
Db        119 GACAGCTTCGGGGGCGAC---GCCCAGGCCGACGAGGGGCAGGCCCGCAAATCGCAGCTG 175
```

```
Qy        189 LeuArgAlaIleArgArgPheLeuHisAsnTyrArg------SerSerAlaHisAspLeu 206
Db        176 CAGAGGCGCTTCAAGGAGTTCCTGCGGCAGTACCGAGTGGGCACCGACCGCACGGGCTTC 235

Qy        207 AsnSerLysTyrIleGlnIleIleGluGluThrValGluArgGluGluAspThrLeuAsn 226
Db        236 ACCTTCAAATACAGGGATGAACTCAAGCGGCATTACAACCTGGGGGAGTACTGGATTGAG 295

Qy        227 IleAspMetSerAspIleTyrAspHisAspProAspLeuTyrAlaLysIleValArgTyr 246
Db        296 GTGGAGATGGAGGATCTGGCCAGCTTTGATGAGGACCTGGCCGACTACTTGTACAAGCAG 355

Qy        247 ProLeuAspIleIleProLeuLeuAspThrGluCysGlnGluValAla---------Thr 263
Db        356 CCAGCCGAGCACCTGCAGCTGCTGGAGGAAGCTGCCAAGGAGGTAGCTGATGAGGTGACC 415

Qy        264 SerLeuLeuProThrPheGluLysHisIleGluAlaArgProPheAsnLeuLysAlaSer 283
Db        416 CGGCCCCGGCCTTCTGGGGAGGAGGTGCTCCAGGACATCCAGGTCATGCTCAAGTCGGAC 475

Qy        284 Val---------HisMetArgGluLeuAsnProSerAspIleAspLysLeuValSerVal 300
Db        476 GCCAGCCCTTCCAGCATTCGTAGCCTGAAGTCGGACATGATGTCACACCTGGTGAAGATC 535

Qy        301 LysGlyMetValIleArgCysSerSerIleIleProGluIleLysGlyAlaPhePheLys 320
Db        536 CCTGGCATCATCATCGCGGCCTCTGCGGTCCGTGCCAAGGCCACCCGCATCTCTATCCAG 595

Qy        321 CysLeuValCysGlyHisSerProProLeuValThrValValLysGly------ArgVal 338
Db        596 TGCCGCAGCTGCCGCAACACCCTCACCAACATTGCCATGCGCCCTGGCCTCGAGGGCTAT 655

Qy        339 GluGluProThrArgCysGlu---------------LysProGluCysAlaAlaArgAsn 353
Db        656 GCCCTGCCCAGGAAGTGCAACACAGATCAGGCTGGGCGCCCCAAATGCCCATTG---GAC 712

Qy        354 AlaMetSerLeuIleHisAsnArgCysThrPheAlaAsnLysGlnIleValArgLeuGln 373
Db        713 CCGTACTTCATCATGCCCGACAAATGCAAATGCGTGGACTTCCAGACCCTGAAGCTGCAG 772

Qy        374 GluThrProAspAlaIleProGluGlyGluThrProHisThrValSerMetCysLeuTyr 393
Db        773 GAGCTGCCTGATGCAGTCCCCCACGGGGAGATGCCCAGACACATGCAGCTCTACTGCGAC 832

Qy        394 AsnThrMetValAspAlaValLysProGlyAspArgIleGluValThrGlyValPheLys 413
Db        833 AGGTACCTGTGTGACAAGGTCGTCCCTGGGAACAGGGTTACCATCATGGGCATCTACTCC 892

Qy        414 AlaMetAlaVal--------------------------ArgValGlyProAsnGlnArg 424
Db        893 ATCAAGAAGTTTGGCCTGACTACCAGCAGGGGCCGTGACAGGGTGGGCGTGGGCATCCGA 952

Qy        425 ThrLeuArgAlaLeuTyrLysThrTyrIleAspCysValHisValLysLysSerAspArg 444
Db        953 AGC-----------------TCCTACATCCGTGTCCTGGGCATCCAG------------ 982

Qy        445 GlyArgLeuGlnThrGluAspProMetGluMetAspLysGluAsnAspMetTyrAlaGly 464
Db        983 ------------------------GTGGACACAGATGGCTCTGGCCGCAGCTTTGCTGGG 1018

Qy        465 TyrHisGluSerAspThrSerGluAlaAlaAsnGluAlaLysIleGlnLysLeuLysGlu 484
Db       1019 ------------------------GCCGTGAGCCCCCAGGAGGAGGAGGAGTTCCGTCGC 1054

Qy        485 LeuSerLysLeuProGlyIleTyrAspArgLeuSerArgSerLeuAlaProSerIleTrp 504
Db       1055 CTGGCTGCCCTCCCAAATGTCTATGAGGTCATCTCCAAGAGCATCGCCCCCTCCATCTTT 1114

Qy        505 GluLeuGluAspIleLysLysGlyLeuLeuCysGlnLeuPheGlyGlyLysAlaLysLys 524
Db       1115 GGGGGCACAGACATGAAGAAGGCCATTGCCTGCCTGCTCTTTGGGGGCTCCCGAAAGAGG 1174

Qy        525 IleProSerGlyAlaSerPheArgGlyAspIleAsnValLeuLeuValGlyAspProGly 544
Db       1175 CTCCCTGATGGACTTACTCGCCGAGGAGACATCAACCTGCTGATGCTAGGGGACCCTGGG 1234

Qy        545 ThrSerLysSerGlnLeuLeuGlnTyrValHisLysIleAlaProArgGlyIleTyrThr 564
Db       1235 ACAGCCAAGTCCCAGCTTCTGAAGTTTGTGGAGAAGTGTTCTCCCATTGGGGTATACACG 1294

Qy        565 SerGlyArgGlySerSerAlaValGlyLeuThrAlaTyrValThrLysAspProGluThr 584
Db       1295 TCTGGGAAAGGCAGCAGCGCAGCTGGACTGACAGCCTCGGTGATGAGGGACCCTTCGTCC 1354

Qy        585 ArgGluThrValLeuGluSerGlyAlaLeuValLeuSerAspArgGlyIleCysCysIle 604
Db       1355 CGGAATTTCATCATGGAGGGCGGAGCCATGGTCCTGGCCGATGGTGGGGTCGTCTGTATT 1414

Qy        605 AspGluPheAspLysMetSerAspAsnAlaArgSerMetLeuHisGluValMetGluGln 624
Db       1415 GACGAGTTTGACAAGATGCGAGAAGATGACCGTGTGGCAATCCACGAAGCCATGGAGCAG 1474

Qy        625 GlnThrValSerValAlaLysGlyGlyIleIleAlaSerLeuAsnAlaArgThrSerVal 644
Db       1475 CAGACCATCTCTATCGCCAAGGCTGGGATCACCACCACCCTGAACTCCCGCTGCTCCGTC 1534

Qy        645 LeuAlaCysAlaAsnProSerGlySerArgTyrAsnAlaArgLeuSerValIleAspAsn 664
Db       1535 CTGGCTGCTGCCAACTCAGTGTTCGGCCGCTGGGATGAGACGAAGGGGGAG---GACAAC 1591

Qy        665 IleGlnLeuProProThrLeuLeuSerArgPheAspLeuIleTyrLeuMetLeuAspLys 684
Db       1592 ATTGACTTCATGCCCACCATCTTGTCGCGCTTCGACATGATCTTCATCGTCAAGGATGAG 1651

Qy        685 ProAspGluGlnAsnAspArgArgLeuAlaArgHisLeuValAlaLeuHisTyrGluAsn 704
Db       1652 CACAATGAGGAGAGGGATGTGATGCTGGCCAAGCATGTCATCACTCTGCACGTGAGCGCA 1711

Qy        705 TyrGluValSerLys------GlnAspAlaLeuAspLeuGlnThrLeuThrAlaTyrIle 722
Db       1712 CTGACACAGACACAGGCTGTGGAGGGCGAGATTGACCTGGCCAAGCTGAAGAAGTTTATT 1771

Qy        723 ThrTyrAlaArgGlnHisValHisProThrLeuSerAspGluAlaAlaGluAspLeuIle 742
Db       1772 GCCTACTGCCGAGTGAAGTGTGGCCCCCGGCTGTCAGCAGAGGCTGCAGAGAAACTGAAG 1831

Qy        743 AsnGlyTyrValGluMetArg---------GlnLysGlyAsnPheProGlySerSerLys 759
Db       1832 AACCGCTACATCATCATGCGGAGCGGGGCCCGTCAGCACGAGAGGGACAGTGACCGCCGC 1891

Qy        760 LysValIleThrAlaThrProArgGlnLeuGluSerMetIleArgIleSerGluAlaLeu 779
Db       1892 TCCAGCATCCCCATCACTGTGCGGCAGCTGGAGGCCATTGTGCGCATCGCGGAAGCCCTC 1951

Qy        780 AlaArgMetArgPheSerGluValValGluLysValAspAlaAlaGluAlaValArgLeu 799
Db       1952 AGCAAGATGAAGCTGCAGCCCTTCGCCACAGAGGCAGATGTGGAGGAGGCCCTGCGGCTC 2011

Qy        800 LeuAspValAlaLeuGlnGlnSerAlaThrAspHisAlaThrGlyThrIleAspMetAsp 819
Db       2012 TTCCAAGTGTCCACGTTGGATGCTGCC---------TTGTCCGGTACCCTG--------- 2053

Qy        820 LeuIleThrThrGlyValSerAlaSerGluArgIleArgArgAlaAsnLeuLeuAlaAla 839
Db       2054 ---------TCAGGGGTGGAGGGCTTCACCAGCCAGGAGGACCAGGAGATGCTGAGCCGC 2104

Qy        840 LeuArgGluLeuIleAlaAspLysIleSerProGlySerSerSerGlyLeuLysThrSer 859
Db       2105 ATCGAGAAGCAGCTCAAGCGCCGCTTTGCCATTGGCTCCCAG------GTGTCTGAGCAC 2158

Qy        860 GlnLeuLeuGluAspIleArgSerGlnSerSerValAspValSerLeuGlnAspIleLys 879
Db       2159 AGCATCATCAAGGACTTCACCAAGCAGAAATACCCGGAGCACGCCATCCACAAGGTGCTG 2218

Qy        880 AsnAlaLeu---------GlySerLeuGln 886
```

```
Db       2219 CAGCTCATGCTGCGGCGCGGCGAGATCCAG 2248

RESULT 8
US-09-976-594-323
; Sequence 323, Application US/09976594
; Patent No. 6673549
; GENERAL INFORMATION:
;  APPLICANT: Furness, Michael
;  APPLICANT:  Buchbinder, Jenny
;  TITLE OF INVENTION: GENES EXPRESSED IN C3A LIVER CELL CULTURES TREATED WITH STEROIDS
;  FILE REFERENCE: PA-0041 US
;  CURRENT APPLICATION NUMBER: US/09/976,594
;  CURRENT FILING DATE:  2001-10-12
;  PRIOR APPLICATION NUMBER: 60/240,409
;  PRIOR FILING DATE: 2000-10-12
;  NUMBER OF SEQ ID NOS: 1143
;  SOFTWARE: PERL Program
; SEQ ID NO 323
;   LENGTH: 3445
;   TYPE: DNA
;   ORGANISM: Homo sapiens
;   FEATURE:
;   NAME/KEY: misc_feature
;   OTHER INFORMATION: Incyte ID No. 6673549 1723834CB1
US-09-976-594-323

Alignment Scores:
Pred. No.:              1.07e-68      Length:         3445
Score:                  936.00        Matches:        261
Percent Similarity:     46.41%        Conservative:   159
Best Local Similarity:  28.84%        Mismatches:     341
Query Match:            20.31%        Indels:         144
DB:                     3             Gaps:           24

US-10-768-511-8 (1-901) x US-09-976-594-323 (1-3445)

Qy         31 ThrSerProSerPheAspAsnAlaAlaSerProValAlaGlyArgArgAlaValArgGln 50
Db         49 TCATCGGAATCCTTCACCATGGCATCCAGCCCG---GCCCAGCGTCGGCGAGGCAATGAT 105

Qy         51 ThrProThrSerAlaValArgArgArgGlyArgGluThrAspSerAlaArgArgArgArg 70
Db        106 CCTCTCACCTCCAGCCCTGGCCGAAGCTCCCGGCGTACTGATGCCCTC------ACCTCC 159

Qy         71 SerArgSerArgSerLeuGlyAsnSerValTyrSerSerProTyrAspAlaGlyThrPro 90
Db        160 AGCCCTGGCCGTGACCTT------------------CCACCATTTGAGGATGAGTCCGAG 201

Qy         91 GlyThrProGlyThrProValAlaThrProValTyrAlaThrProValGlyThrProMet 110
Db        202 GGGCTCCTAGGCACA------GAGGGGCCCCTGGAGGAAGAAGAGGATGGAGAGGAGCTC 255

Qy        111 ---------GlyThrProSerPheHisArgGlyThrProGlnTyrLysGlnArgSerGlu 127
Db        256 ATTGGAGATGGCATGGAAAGGGACTACCGCGCCATCCCAGAGCTGGACGCCTATGAGGCC 315

Qy        128 LeuGly---------------------------------------SerGlnGlyLysPro 134
Db        316 GAGGGACTGGCTCTGGATGATGAGGACGTAGAGGAGCTGACGGCCAGTCAGAGGGAGGCA 375

Qy        135 LeuHisArgArgArgArgSerGlnSerArgGluProGly--------------------- 147
Db        376 GCAGAGCGGGCCATGCGGCAGCGTGACCGGGAGGCTGGCCGAGGCCTGGGCCGCATGCGC 435

Qy        148 -------------------------------HisArgSerProSerArgGluProSer 156
Db        436 CGTGGGCTCCTGTATGACAGCGATGAGGAGGACGAGGAGCGCCCTGCCCGCAAGCGCCGC 495

Qy        157 AlaAspGlyArgProSerGluSerAlaGluProAsp----------------------- 168
Db        496 CAGGTGGAGCGGGCCACGGAGGACGGCGAGGAGGACGAGGAGATGATTGAGAGCATCGAG 555

Qy        169 ------AspThrLeuGlyGlyGluTyrAlaTyrValTrp-----------GlyThrAsn 182
Db        556 AACCTGGAGGATCTCAAAGGCCACTCTGTGCGCGAGTGGGTGAGCATGGCGGGCCCCCGG 615

Qy        183 ValAsnIleProAspValLeuArgAlaIleArgArgPheLeuHisAsnTyr---ArgSer 201
Db        616 CTG---------GAGATCCACCACCGCTTCAAGAACTTCCTGCGCACTCACGTCGACAGC 666

Qy        202 SerAlaHisAspLeuAsnSerLysTyrIleGlnIleIleGluGluThrValGluArgGlu 221
Db        667 CACGGCCACAACGTC---------TTCAAGGAGCGTATCAGCGACATGTGCAAAGAGAAC 717

Qy        222 GluAspThrLeuAsnIleAspMetSerAspIleTyrAspHisAspProAspLeuTyrAla 241
Db        718 CGTGAGAGCCTGGTGGTGAACTATGAGGACTTGGCAGCCAGGGAGCACGTGCTGGCCTAC 777

Qy        242 LysIleValArgTyrProLeuAspIleIleProLeuLeuAspThrGluCysGlnGluVal 261
Db        778 TTCCTGCCTGAGGCACCGGCGGAGCTGCTGCAGATCTTTGATGAGGCTGCCCTGGAGGTG 837

Qy        262 AlaThrSerLeuLeuProThrPheGluLys---------HisIleGluAlaArgProPhe 278
Db        838 GTACTGGCCATGTACCCCAAGTACGACCGCATCACCAACCACATCCATGTCCGCATCTCC 897

Qy        279 AsnLeuLysAlaSerValHisMetArgGluLeuAsnProSerAsp--------------- 293
Db        898 CACCTGCCTCTGGTGGAGGAGCTGCGCTCGCTGAGGCGAGCTGAGGGCAGGCAGCTGCAT 957

Qy        294 IleAspLysLeuValSerValLysGlyMetValIleArgCysSerSerIleIleProGlu 313
Db        958 CTGAACCAGCTGATCCGCACCAGTGGGGTGGTGACCAGCTGCACTGGCGTCCTGCCCCAG 1017

Qy        314 IleLysGlyAlaPhePheLysCysLeuValCysGlyHisSerProProLeuValThrVal 333
Db       1018 CTCAGCATGGTCAAGTACAACTGCAACAAGTGCAATTTCGTCCTGGGTCCTTTCTGCCAG 1077

Qy        334 ValLysGlyArgValGluGluProThrArgCysGluLysProGluCysAlaAlaArgAsn 353
Db       1078 TCCCAGAACCAGGAGGTGAAACCAGGCTCCTGT------CCTGAGTGCCAGTCGGCCGGC 1131

Qy        354 AlaMetSerLeuIleHisAsnArgCysThrPheAlaAsnLysGlnIleValArgLeuGln 373
Db       1132 CCCTTTGAGGTCAACATGGAGGAGACCATCTATCAGAACTACCAGCGTATCCGAATCCAG 1191

Qy        374 GluThrProAspAlaIleProGluGlyGluThrProHisThrValSerMetCysLeuTyr 393
Db       1192 GAGAGTCCAGGCAAAGTGGCGGCTGGCCGGCTGCCCCGCTCCAAGGACGCCATTCTCCTC 1251

Qy        394 AsnThrMetValAspAlaValLysProGlyAspArgIleGluValThrGlyValPheLys 413
Db       1252 GCAGATCTGGTGGACAGCTGCAAGCCAGGAGACGAGATAGAGCTGACTGGCATCTATCAC 1311

Qy        414 AlaMetAlaValArgValGlyProAsnGlnArgThrLeuArgAla------LeuTyrLys 431
Db       1312 AAC---------AACTATGATGGCTCCCTCAACACTGCCAATGGCTTCCCTGTCTTTGCC 1362

Qy        432 ThrTyrIleAspCysValHisValLysLysSerAspArg-----------GlyArgLeu 447
Db       1363 ACTGTCATCCTAGCCAACCACGTGGCCAAGAAGGACAACAAGGTTGCTGTAGGGGAACTG 1422

Qy        448 GlnThrGluAspProMetGluMetAspLysGluAsnAspMetTyrAlaGlyTyrHisGlu 467
Db       1423 ACCGATGAAGAT------------------------------------------------ 1434

Qy        468 SerAspThrSerGluAlaAlaAsnGluAlaLysIleGlnLysLeuLysGluLeuSerLys 487
Db       1435 ---------------------------------GTGAAGATGATCACTAGCCTCTCCAAG 1461

Qy        488 LeuProGlyIleTyrAspArgLeuSerArgSerLeuAlaProSerIleTrpGluLeuGlu 507
Db       1462 GATCAGCAGATCGGAGAGAAGATCTTTGCCAGCATTGCTCCTTCCATCTATGGTCATGAA 1521

Qy        508 AspIleLysLysGlyLeuLeuCysGlnLeuPheGlyGlyLysAlaLysLysIleProSer 527
```

```
Db     1522 GACATCAAGAGAGGCCTGGCTCTGGCCCTGTTCGGAGGGGAGCCCAAAAACCCAGGTGGC 1581

Qy      528 GlyAlaSerPheArgGlyAspIleAsnValLeuLeuValGlyAspProGlyThrSerLys 547

Db     1582 AAGCACAAGGTACGTGGTGATATCAACGTGCTCTTGTGCGGAGACCCTGGCACAGCGAAG 1641

Qy      548 SerGlnLeuLeuGlnTyrValHisLysIleAlaProArgGlyIleTyrThrSerGlyArg 567

Db     1642 TCGCAGTTTCTCAAGTATATTGAGAAAGTGTCCAGCCGAGCCATCTTCACCACTGGCCAG 1701

Qy      568 GlySerSerAlaValGlyLeuThrAlaTyrValThrLysAspProGluThrArgGluThr 587

Db     1702 GGGGCGTCGGCTGTGGGCCTCACGGCGTATGTCCAGCGGCACCCTGTCAGCAGGGAGTGG 1761

Qy      588 ValLeuGluSerGlyAlaLeuValLeuSerAspArgGlyIleCysCysIleAspGluPhe 607

Db     1762 ACCTTGGAGGCTGGGGCCCTGGTTCTGGCTGACCGAGGAGTGTGTCTCATTGATGAATTT 1821

Qy      608 AspLysMetSerAspAsnAlaArgSerMetLeuHisGluValMetGluGlnGlnThrVal 627

Db     1822 GACAAGATGAATGACCAGGACAGAACCAGCATCCATGAGGCCATGGAGCAACAGAGCATC 1881

Qy      628 SerValAlaLysGlyGlyIleIleAlaSerLeuAsnAlaArgThrSerValLeuAlaCys 647

Db     1882 TCCATCTCGAAGGCTGGCATCGTCACCTCCCTGCAGGCTCGCTGCACGGTCATTGCTGCC 1941

Qy      648 AlaAsnProSerGlySerArgTyrAsnAlaArgLeuSerValIleAspAsnIleGlnLeu 667

Db     1942 GCCAACCCCATAGGAGGGCGCTACGACCCCTCGCTGACTTTCTCTGAGAACGTGGACCTC 2001

Qy      668 ProProThrLeuLeuSerArgPheAspLeuIleTyrLeuMetLeuAspLysProAspGlu 687

Db     2002 ACAGAGCCCATCATCTCACGCTTTGACATCCTGTGTGTGGTGAGGGACACCGTGGACCCA 2061

Qy      688 GlnAsnAspArgArgLeuAlaArgHisLeuValAlaLeuHisTyrGluAsnTyrGluVal 707

Db     2062 GTCCAGGACGAGATGCTGGCCCGCTTCGTGGTGGGCAGCCACGTCAGACACCACCCCAGC 2121

Qy      708 SerLysGln--------------------------------------------------- 710

Db     2122 AACAAGGAGGAGGAGGGGCTGGCCAATGGCAGCGCTGCTGAGCCCGCCATGCCCAACACG 2181

Qy      711 ---------AspAlaLeuAspLeuGlnThrLeuThrAlaTyrIleThrTyrAlaArgGln 727

Db     2182 TATGGCGTGGAGCCCCTGCCCAGGAGGTCCTGAAGAAGTACATCATCTACGCCAAGGAG 2241

Qy      728 HisValHisProThrLeuSerAspGluAlaAlaGluAspLeuIleAsnGlyTyrValGlu 747

Db     2242 AGGGTCCACCCGAAGCTCAACCAGATGGACCAGGACAAGGTGGCCAAGATGTACAGTGAC 2301

Qy      748 MetArgGlnLysGlyAsnPheProGlySerSerLysLysValIleThrAlaThrProArg 767

Db     2302 CTGAGGAAAGAATCTATGGCGACAGGCAGC------------ATCCCCATTACGGTGCGG 2349

Qy      768 GlnLeuGluSerMetIleArgIleSerGluAlaLeuAlaArgMetArgPheSerGluVal 787

Db     2350 CACATCGAGTCCATGATCCGCATGGCGGAGGCCCACGCGCGCATCCATCTGCGGGACTAT 2409

Qy      788 ValGluLysValAspAlaAlaGluAlaValArgLeuLeuAspValAlaLeuGlnGlnSer 807

Db     2410 GTGATCGAAGACGACGTCAACATGGCCATCCGCGTG------------ATGCTGGAGAGC 2457

Qy      808 AlaThrAspHisAlaThrGlyThrIleAspMetAspLeuIleThrThrGlyValSerAla 827

Db     2458 TTCATAGACACACAGAAGTTCAGCGTCATGCGCAGCATGCGCAAGACT---TTTGCCCGC 2514

Qy      828 SerGluArgIleArgArgAlaAsn---------LeuLeuAlaAlaLeuArgGluLeuIle 844

Db     2515 TACCTTTCATTCCGGCGTGACAACAATGAGCTGTTGCTCTTCATACTGAAGCAGTTAGTG 2574

Qy      845 AlaAspLysIleSer 849

Db     2575 GCAGAGCAGGTGACA 2589
```

```
RESULT 9
US-09-949-016-2468
; Sequence 2468, Application US/09949016
; Patent No. 6812339
; GENERAL INFORMATION:
;  APPLICANT: VENTER, J. Craig et al.
;  TITLE OF INVENTION: POLYMORPHISMS IN KNOWN GENES ASSOCIATED
;  TITLE OF INVENTION:  WITH HUMAN DISEASE, METHODS OF DETECTION AND USES THEREOF
;  FILE REFERENCE: CL001307
;  CURRENT APPLICATION NUMBER: US/09/949,016
;  CURRENT FILING DATE:  2000-04-14
;  PRIOR APPLICATION NUMBER: 60/241,755
;  PRIOR FILING DATE: 2000-10-20
;  PRIOR APPLICATION NUMBER: 60/237,768
;  PRIOR FILING DATE: 2000-10-03
;  PRIOR APPLICATION NUMBER: 60/231,498
;  PRIOR FILING DATE: 2000-09-08
;  NUMBER OF SEQ ID NOS: 207012
;  SOFTWARE: FastSEQ for Windows Version 4.0
; SEQ ID NO 2468
;   LENGTH: 2917
;   TYPE: DNA
;   ORGANISM: Human
US-09-949-016-2468

Alignment Scores:
Pred. No.:               2.46e-67      Length:         2917
Score:                   918.50        Matches:        267
Percent Similarity:      45.43%        Conservative:   121
Best Local Similarity:   31.26%        Mismatches:     301
Query Match:             19.93%        Indels:         165
DB:                      3             Gaps:           24

US-10-768-511-8 (1-901) x US-09-949-016-2468 (1-2917)

Qy      124 GlnArgSerGluLeuGlySerGlnGlyLysProLeuHisArgArgArgArgSerGlnSer 143

Db       33 CGCAGGCGAATCCTCGGCACTAAGCAAATATGGACCTCGCGGCGGCAGCGGAGCCGGGCG 92

Qy      144 -ArgGluProGlyHisArgSerProSerArgGluProSerAlaAspGlyArgProSerGl 163

Db       93 CCGGCAGCCAGCACCTGGAGGTCCGCGACGA------------GGTGGCCGAGAAGTGCC 140

Qy      163 uSerAlaGluPro-AspAspThrLeuGlyGlyGluTyrAlaTyrValTrpGlyThrAsnV 183

Db      141 AGAAACTGTTCCTGGAC------------------------------------------- 157

Qy      183 alAsnIleProAspValLeuArgAlaIleArgArgPheLeuHisAsnTyrArgSerSerA 203

Db      158 -----------------------------------TTCTTGGAGGAGTTTCAGAGCAGC- 181

Qy      203 laHisAspLeuAsnSerLysTyrIleGlnIleIleGluGluThrValGluArgGluGluA 223

Db      182 -----GATGGAGAAATTAAATACTTGCAATTAGCAGAGGAACTGATTCGTCCTGAGAGAA 236

Qy      223 spThrLeuAsnIleAspMetSerAspIleTyrAspHisAspProAspLeuTyrAlaLysI 243

Db      237 ACACATTGGTTGTGAGTTTTGTGGACCTGGAACAATTTAACCAGCAACTTTCCACCACCA 296

Qy      243 leValArgTyrProLeuAspIleIleProLeuLeuAspThrGluCysGlnGluValAlaT 263

Db      297 TTCAAGAGGAGTTCTATAGAGTTTACCCTTACCTG---------TGTCGG---------- 337

Qy      263 hrSerLeuLeuProThrPheGluLysHisIleGluAlaArgProPhe------------- 278

Db      338 -----GCCTTGAAAACATTCGTCAAAGACCGTAAAGAGATCCCTCTTGCCAAGGATTTTT 392

Qy      279 --------------AsnLeuLysAlaSerValHisMetArgGluLeuAsnProSerAspI 294

Db      393 ATGTTGCATTCCAAGACCTGCCTACCAGACACAAGATTCGAGAGCTCACCTCATCCAGAA 452

Qy      294 leAspLysLeuValSerValLysGlyMetValIleArgCysSerSerIleIleProGluI 314
```

```
Db        453 TTGGTTTGCTCACTCGCATCAGTGGGCAGGTGGTGCGGACTCACCCAGTTCACCCAGAGC 512

Qy        314 leLysGlyAlaPhePheLysCysLeuValCysGlyHisSerProProLeuValThrValV 334

Db        513 TTGTGAGCGGAACTTTTCTGTGCTTGGACTGT------------------CAGACAGTGA 554

Qy        334 alLysGly---------------ArgValGluGluProThrArgCysGluLysProGluC 349

Db        555 TCAGGGATGTAGAACAGCAGTTCAAATACACACAGCCAAACATCTGCCGAAATCCAGTTT 614

Qy        349 ysAlaAlaArgAsnAlaMetSerLeuIleHisAsnArgCysThrPheAlaAsnLysGlnI 369

Db        615 GTGCCAACAGGAGGAGATTCTTACTGGATACAAATAAATCAAGATTTGTTGATTTTCAAA 674

Qy        369 leValArgLeuGlnGluThrProAspAlaIleProGluGlyGluThrProHisThrValS 389

Db        675 AGGTTCGTATTCAAGAGACCCAAGCTGAGCTTCCTCGAGGGAGTATCCCCCGCAGTTTAG 734

Qy        389 erMetCysLeuTyrAsnThrMetValAspAlaValLysProGlyAspArgIleGluValT 409

Db        735 AAGTAATTTTAAGGGCTGAAGCTGTGGAATCAGCTCAAGCTGGTGACAAGTGTGACTTTA 794

Qy        409 hrGlyVal---------------------------PheLysAlaMetAlaValArgValG 420

Db        795 CAGGGACACTGATTGTTGTGCCTGACGTCTCCAAGCTTAGCACACCAGGAGCACGTGCAG 854

Qy        420 lyProAsnGlnArgThr---------------------------------------LeuA 427

Db        855 AAACTAATTCCCGTGTCAGTGGTGTTGATGGATATGAGACAGAAGGCATTCGAGGACTCC 914

Qy        427 rgAlaLeu------------------TyrLys---ThrTyrIleAspCysValHisValL 440

Db        915 GGGCCCTTGGTGTTAGGGACCTTTCTTATAGGCTGGTCTTTCTTGCCTGCTGT------- 967

Qy        440 ysLysSerAspArgGlyArgLeuGlnThrGluAspProMetGluMetAspLysGluAsnA 460

Db        968 --------------------------------------------GTTGCGCCAACCAACC 983

Qy        460 spMetTyrAlaGlyTyrHisGluSerAspThrSerGluAlaAlaAsnGluAlaLysIleG 480

Db        984 CAAGGTTTGGGGGGAAAGAGCTCAGAGATGAGGAACAGACAGCTGAGAGCATTAAGAACC 1043

Qy        480 ln--------------------LysLeuLysGluLeuSerLysLeuProGlyIleTyrA 493

Db       1044 AAATGACTGTGAAAGAATGGGAGAAAGTGTTTGAGATGAGTCAAGATAAAAATCTATACC 1103

Qy        493 spArgLeuSerArgSerLeuAlaProSerIleTrpGluLeuGluAspIleLysLysGlyL 513

Db       1104 ACAATCTTTGTACCAGCCTGTTCCCTACTATACATGGCAATGATGAAGTAAAACGGGGTG 1163

Qy        513 euLeuCysGlnLeuPheGlyGlyLysAlaLysLysIleProSerGlyAlaSerPheArgG 533

Db       1164 TCCTGCTGATGCTCTTTGGTGGCGTTCCAAAGACAACAGGAGAAGGACCTCTCTTCGAG 1223

Qy        533 lyAspIleAsnValLeuLeuValGlyAspProGlyThrSerLysSerGlnLeuLeuGlnT 553

Db       1224 GGGACATAAATGTTTGCATTGTTGGTGACCCAAGTACAGCTAAGAGCCAATTTCTCAAGC 1283

Qy        553 yrValHisLysIleAlaProArgGlyIleTyrThrSerGlyArgGlySerSerAlaValG 573

Db       1284 ACGTGGAGGAGTTCAGCCCCAGAGCTGTCTACACCAGTGGTAAAGCGTCCAGTGCTGCTG 1343

Qy        573 lyLeuThrAlaTyrValThrLysAspProGluThrArgGluThrValLeuGluSerGlyA 593

Db       1344 GCTTAACAGCAGCTGTTGTGAGAGATGAAGAATCTCATGAGTTTGTCATTGAGGCTGGAG 1403

Qy        593 laLeuValLeuSerAspArgGlyIleCysCysIleAspGluPheAspLysMetSerAspA 613

Db       1404 CTTTGATGTTGGCTGATAATGGTGTGTGTTGTATTGATGAATTTGATAAGATGGACGTGC 1463

Qy        613 snAlaArgSerMetLeuHisGluValMetGluGlnGlnThrValSerValAlaLysGlyG 633

Db       1464 GGGATCAAGTTGCTATTCATGAAGCTATGGAACAGCAGACCATATCCATCACTAAAGCAG 1523

Qy        633 lyIleIleAlaSerLeuAsnAlaArgThrSerValLeuAlaCysAlaAsnProSerGlyS 653

Db       1524 GAGTGAAGGCTACTCTGAACGCCCGGACGTCCATTTTGGCAGCAGCAAACCCAATCAGTG 1583

Qy        653 erArgTyrAsnAlaArgLeuSerValIleAspAsnIleGlnLeuProProThrLeuLeuS 673

Db       1584 GACACTATGACAGATCAAAATCATTGAAACAGAATATAAATTTGTCAGCTCCCATCATGT 1643

Qy        673 erArgPheAspLeuIleTyrLeuMetLeuAspLysProAspGluGlnAsnAspArgArgL 693

Db       1644 CCCGATTCGATCTCTTCTTTATCCTTGTGGATGAATGTAATGAGGTTACAGATTATGCCA 1703

Qy        693 euAlaArgHisLeuValAlaLeuHisTyrGluAsnTyrGluValSerLysGlnAspAlaL 713

Db       1704 TTGCCAGGCGCATAGTAGATTTGCAT---TCAAGAATTGAGGAATCAATTGATCGTGTCT 1760

Qy        713 euAspLeuGlnThrLeuThrAlaTyrIleThrTyrAlaArgGlnHisValHisProThrL 733

Db       1761 ATTCCCTCGATGATATCAGAAGATATCTTCTCTTTGCAAGACAG---TTTAAACCCAAGA 1817

Qy        733 euSerAspGluAlaAlaGluAspLeuIleAsnGlyTyrValGluMetArgGlnLysGlyA 753

Db       1818 TTTCCAAAGAGTCAGAGGACTTCATTGTGGAGCAATATAAACATCTCCGCCAGAGAGATG 1877

Qy        753 snPheProGlySerSerLysLysValIleThrAlaThrProArgGlnLeuGluSerMetI 773

Db       1878 GT---TCTGGAGTGACCAAGTCTTCATGGAGGATTACAGTGCGACAGCTTGAGAGCATGA 1934

Qy        773 leArgIleSerGluAlaLeuAlaArgMetArgPheSerGluValValGluLysValAspA 793

Db       1935 TTCGTCTCTCTGAAGCTATGGCTCGGATGCACTGCTGTGATGAGGTCCAACCTAAACATG 1994

Qy        793 laAlaGluAlaValArgLeuLeu-----------------------------AspValA 803

Db       1995 TGAAGGAAGCTTTCCGGTTACTGAATAAATCAATCATCCGTGTGGAAACACCTGATGTCA 2054

Qy        803 laLeuGlnGlnSer---------------AlaThrAspHisAlaThrGlyThrIle---- 816

Db       2055 ATCTAGATCAAGAGGAAGAGATCCAGATGGAGGTAGATGAGGGTGCTGGTGGCATCAATG 2114

Qy        817 --------AspMetAspLeuIleThrThrGlyValSerAla------------------ 827

Db       2115 GTCATGCTGACAGCCCTGCTCCTGTGAACGGGATCAATGGCTACAATGAAGACATAAATC 2174

Qy        828 --------------------------------------SerGluArgIleArgArgAlaA 835

Db       2175 AAGAGTCTGCTCCCAAAGCCTCCTTAAGGCTGGGCTTCTCTGAGTACTGCCGAATCTCTA 2234

Qy        835 snLeuLeuAlaAla---LeuArgGluLeuIleAlaAspLysIleSerProGlySerSerS 854

Db       2235 ACCTTATTGTGCTTCACCTCAGAAAGGTGGAAGAAGAAGAG-----------GACGAGT 2282

Qy        854 erGlyLeuLysThrSerGlnLeu-------------LeuGluAspIleArgSerGlnSerS 870

Db       2283 CAGCATTAAAGAGGAGCGAGCTTGTTAACTGGTACTTGAAGGAAATCGAATCAGAGATAG 2342

Qy        870 erValAspValSerLeuGlnAspIleLysAsnAlaLeu 882

Db       2343 ACTCTGAAGAAGAACTTATAAATAAAAAAAGAATCATA 2380
```

RESULT 10
US-09-220-132-12
; Sequence 12, Application US/09220132
; Patent No. 6506607
; GENERAL INFORMATION:
; APPLICANT: Shyjan, Andrew W.
; TITLE OF INVENTION: METHODS AND COMPOSITONS FOR THE IDENTIFICATION AND ASSESSMENT
; TITLE OF INVENTION: OF PROSTATE CANCER THERAPIES AND THE DIAGNOSIS OF PROSTATE CANC
; FILE REFERENCE: 07334-074001
; CURRENT APPLICATION NUMBER: US/09/220,132
; CURRENT FILING DATE: 1998-12-23

```
;  PRIOR APPLICATION NUMBER:  US 60/079,303
;  PRIOR FILING DATE:  1998-03-25
;  PRIOR APPLICATION NUMBER:  US 60/068,821
;  PRIOR FILING DATE:  1997-12-24
;  NUMBER OF SEQ ID NOS:  191
;  SOFTWARE:  FastSEQ for Windows Version 4.0
; SEQ ID NO 12
;   LENGTH: 3379
;   TYPE: DNA
;   ORGANISM: Homo sapiens
US-09-220-132-12

Alignment Scores:
Pred. No.:               6.03e-62      Length:         3379
Score:                   856.00        Matches:        244
Percent Similarity:      45.26%        Conservative:   143
Best Local Similarity:   28.54%        Mismatches:     322
Query Match:             18.57%        Indels:         147
DB:                      3             Gaps:           19

US-10-768-511-8 (1-901) x US-09-220-132-12 (1-3379)

Qy     31 ThrSerProSerPheAspAsnAlaAlaSerProValAlaGlyArgArgAlaValArgGln 50
Db     13 TCATCGGAATCCTTCACCATGGCATCCAGCCCG---GCCCAGCGTCGGCGAGGCAATGAT 69

Qy     51 ThrProThrSerAlaValArgArgArgGlyArgGluThrAspSerAlaArgArgArgArg 70
Db     70 CCTCTCACCTCCAGCCCTGGCCGAAGCTCCCGGCGTACTGATGCCCTC------ACCTCC 123

Qy     71 SerArgSerArgSerLeuGlyAsnSerValTyrSerSerProTyrAspAlaGlyThrPro 90
Db    124 AGCCCTGGCCGTGACCTT------------------CCACCATTTGAGGATGAGTCCGAG 165

Qy     91 GlyThrProGlyThrProValAlaThrProValTyrAlaThrProValGlyThrProMet 110
Db    166 GGGCTCCTAGGCACA---------------------GAGGGGCCCCT-GGAGGAAGAAGA 203

Qy    111 GlyThrProSerPheHis---------------ArgGlyThrProGlnTyrLysGlnArg 125
Db    204 GGATGGAGAGGAGCTCATTGGAGATGGCATGGAAAGGGACTACCGCGCCATCCCAGAGCT 263

Qy    126 SerGluLeu--------------------------------------------------- 128
Db    264 GGACGCCTATGAGGCCGAGGGACTGGCTCTGGATGATGAGGACGTAGAGGAGCTGACGGC 323

Qy    129 ------------GlySerGlnGlyLysProLeuHis------------------------ 136
Db    324 CAGTCGAAGGGAGGCAGCAGACGGGCCATGCGGCACGGTGACCGGGAGCTGGCCGGGGCT 383

Qy    136 ------------------------------------------------------------ 136
Db    384 GGGCGCATGCGCCGTGGGCTCCTGTATGACAGCGATGAGGAGGACGAGGAGCGCCCTGCC 443

Qy    137 ArgArgArgArgSerGlnSerArgGluProGlyHisArgSerProSerArgGluProSer 156
Db    444 CGCAAGCGCCGCCAGTGGAGCCGGCACGGAGGACGGCGAGGAGGACGAGCAGATGATTGA 503

Qy    157 AlaAspGlyArgProSerGluSerAlaGluProAspAspThrLeuGlyGlyGluTyrAla 176
Db    504 GAGCATCGAGAACCTGGAGGATCTCAAAGGCCACTCTGTGCGCGAGTGGGTGAGCATGGC 563

Qy    177 TyrValTrpGlyThrAsnValAsnIleProAspValLeuArgAlaIleArgArgPheLeu 196
Db    564 ---------GGGCCCCCGGCTGGAGATCCACCACCGCTTCAAGAACTTCCT------GCG 608

Qy    197 HisAsnTyrArg-SerSerAlaHisAspLeuAsnSerLysTyrIleGlnIleIleGluGl 216
Db    609 CACTCACGTCGACAGCCACGGCCACAACGTC---------TTCAAGGAGCGCATCAGCGA 659

Qy    216 uThrValGluArgGluGluAspThrLeuAsnIleAspMetSerAspIleTyrAspHisAs 236
Db    660 CATGTGCAAAGAGAACCGTGAGAGCCTGGTGGTGAACTATGACGACTTGGCAGCCAGGGA 719

Qy    236 pProAspLeuTyrAlaLysIleValArgTyrProLeuAspIleIleProLeuLeuAspTh 256
Db    720 GCACGTGCTGGCCTACTTCCTGCCTGAGGCACCGGCGGAGCTGCTGCAGATCTTTGATGA 779

Qy    256 rGluCysGlnGluValAlaThrSerLeuLeuProThrPheGluLys---------HisIl 273
Db    780 GGCTGCCCTGGAGGTGGTACTGGCCATGTACCCCAAGTACGACCGCATCACCAACCACAT 839

Qy    273 eGluAlaArgProPheAsnLeuLysAlaSerValHisMetArgGluLeuAsnProSerAs 293
Db    840 CCATGTCCGCATCTCCCACCTGCCTCTGGTGGAGGAGCTGCGCTCGCTGAGGCAGCTGCA 899

Qy    293 pIleAspLysLeuValSerValLysGlyMetValIleArgCysSerSerIleIleProGl 313
Db    900 TCTGAACCAGCTGATCCGCACCAGTGGGGTGGTGACCAGCTGCACTGGCGTCCTGCCCCA 959

Qy    313 uIleLysGlyAlaPhePheLysCysLeuValCysGlyHisSerProProLeuValThrVa 333
Db    960 GCTCAGCATGGTCAAGTACAACTGCAACAAGTGCAATTTCGTCCTGGGTCCTTTCTGCCA 1019

Qy    333 lValLysGlyArgValGluGluProThrArgCysGluLysProGluCysAlaAlaArgAs 353
Db   1020 GTCCCAGAACCAGGAGGTGAAACCAGGCTCCTGT------CCTGAGTGCCAGTCGGCCGG 1073

Qy    353 nAlaMetSerLeuIleHisAsnArgCysThrPheAlaAsnLysGlnIleValArgLeuGl 373
Db   1074 CCCCTTTGAGGTCAACATGGAGGAGACCATCTATCAGAACTACCAGCGTATCCGAATCCA 1133

Qy    373 nGluThrProAspAlaIleProGluGlyGluThrProHisThrValSerMetCysLeuTy 393
Db   1134 GGAGAGTCCAGGCAAAGTGGCGGCTCGGCGGCTGCCCCGCTCCAAGGACGCCATTCTCCT 1193

Qy    393 rAsnThrMetValAspAlaValLysProGlyAspArgIleGluValThrGlyValPheLy 413
Db   1194 CGCAGATCTGGTGGACAGCTGCAACGCAGGAGACGAGATAGAGCTGACTGGCATCTATCA 1253

Qy    413 sAlaMetAlaValArgValGlyProAsnGlnArgThrLeuArgAla------LeuTyrLy 431
Db   1254 CAAC---------AACTATGATGGCTCCCTCAACACTGCCAATGGCTTCCCTGTCTTTGC 1304

Qy    431 sThrTyrIleAspCysValHisValLysLysSerAspArg------------GlyArgLe 447
Db   1305 CACTGTCATCCTAGCCAACCACGTGGCCAAGAAGGACAACAAGGTTGCTGTAGGGGAACT 1364

Qy    447 uGlnThrGluAspProMetGluMetAspLysGluAsnAspMetTyrAlaGlyTyrHisGl 467
Db   1365 GACCGATGAAGAT----------------------------------------------- 1377

Qy    467 uSerAspThrSerGluAlaAlaAsnGluAlaLysIleGlnLysLeuLysGluLeuSerLy 487
Db   1378 ----------------------------------GTGAAGATGATCACTAGCCTCTCCAA 1403

Qy    487 sLeuProGlyIleTyrAspArgLeuSerArgSerLeuAlaProSerIleTrpGluLeuGl 507
Db   1404 GGATCAGCAGATCGGAGAGAAGATCTTTGCCAGCATTGCTCCTTCCATCTATGGTCATGA 1463

Qy    507 uAspIleLysLysGlyLeuLeuCysGlnLeuPheGlyGlyLysAlaLysLysIleProSe 527
Db   1464 AGACATCAAGAGAGGCCCTGCTCTGGCCCTGTTCGGAGGGGAGCCCAAAAACCCAGGTGG 1523

Qy    527 rGlyAlaSerPheArgGlyAspIleAsnValLeuLeuValGlyAspProGlyThrSerLy 547
Db   1524 CAAGCACAAGGTACGTGGTGATATCAACGTGCTCTTGTGCGGAGACCCTGGCACAGCGAA 1583

Qy    547 sSerGlnLeuLeuGlnTyrValHisLysIleAlaProArgGlyIleTyrThrSerGlyAr 567
Db   1584 GTCGCAGTTTCTCAAGTATATTGAGAAAGTGTCCAGCCGAGCCATCTTCACCACTGGCCA 1643

Qy    567 gGlySerSerAlaValGlyLeuThrAlaTyrValThrLysAspProGluThrArgGluTh 587
Db   1644 GGGGGCGTCGGCTGTGGCCGTCACGGCGTATGTCCAGCGGCACCCTGTCAGCAGGGAGTG 1703
```

```
Qy        587 rValLeuGluSerGlyAlaLeuValLeuSerAspArgGlyIleCysCysIleAspGluPh 607
Db       1704 GACCTTGGAGGCTGGGGCCCTGGTTCTGGCTGACCGAGGAGTGTGTCTCATTGATGAATT 1763

Qy        607 eAspLysMetSerAspAsnAlaArgSerMetLeuHisGluValMetGluGlnGlnThrVa 627
Db       1764 TGACAAGATGAATGACCAGGACAGAACCAGCATCCATGAGGCCATGGAGCAACAGAGCAT 1823

Qy        627 lSerValAlaLysGlyGlyIleIleAlaSerLeuAsnAlaArgThrSerValLeuAlaCy 647
Db       1824 CTCCATCTCGAAGGCTGGCATCGTCACCTCCCTGCAGGCTCGCTGCACGGTCATTGCTGC 1883

Qy        647 sAlaAsnProSerGlySerArgTyrAsnAlaArgLeuSerValIleAspAsnIleGlnLe 667
Db       1884 CGCCAACCCCATAGGAGGGCGCTACGACCCCTCGCTGACTTTCTCTGAGAACGTGGACCT 1943

Qy        667 uProProThrLeuLeuSerArgPheAspLeuIleTyrLeuMetLeuAspLysProAspGl 687
Db       1944 CACAGAGCCCATCATCTCACGCTTTGACATCCTGTGTGTGGTGAGGGACACCGTGGACCC 2003

Qy        687 uGlnAsnAspArgArgLeuAlaArgHisLeuValAlaLeuHisTyrGluAsnTyrGluVa 707
Db       2004 AGTCCAGGACGAGATGCTGGCCCGCTTCGTGGTGGGCAGCCACGTCAGACACCACCCCAG 2063

Qy        707 lSerLysGln-------------------------------------------------- 710
Db       2064 CAACAAGGAGGAGGAGGGGCTGGCCAATGGCAGCGCTGCTGAGCCCGCCATGCCCAACAC 2123

Qy        711 ----------AspAlaLeuAspLeuGlnThrLeuThrAlaTyrIleThrTyrAlaArgGl 727
Db       2124 GTATGGCGTGGAGCCCCTGCCCCAGGAGGTCCTGAAGAAGTACATCATCTACGCCAAGGA 2183

Qy        727 nHisValHisProThrLeuSerAspGluAlaAlaGluAspLeuIleAsnGlyTyrValGl 747
Db       2184 GAGGGTCCACCCGAAGCTCAACCAGATGGACCAGGACAAGGTGGCCAAGATGTACAGTGA 2243

Qy        747 uMetArgGlnLysGlyAsnPheProGlySerSerLysLysValIleThrAlaThrProAr 767
Db       2244 CCTGAGGAAAGAATCTATGGCGACAGGCAGC------------ATCCCCATTACGGTGCG 2291

Qy        767 gGlnLeuGluSerMet-IleArgIleSerGluAlaLeuAlaArgMetArgPheSerGluV 787
Db       2292 GCACATCGAGTCCATGAGTC---ATGGCGGAGGCCCACGCGCGCATCCATCTGCGGGACT 2348

Qy        787 alValGluLysValAspAlaAlaGluAlaValArgLeuLeu 800
Db       2349 ATGTGATCGAAGACGACGTCAACATGGCCATCCGCGTGATG 2389
```

RESULT 11
US-09-248-796A-4534
; Sequence 4534, Application US/09248796A
; Patent No. 6747137
; GENERAL INFORMATION:
; APPLICANT: Keith Weinstock et al
; TITLE OF INVENTION: NUCLEIC ACID AND AMINO ACID SEQUENCES RELATING TO CANDIDA ALBICAN
; TITLE OF INVENTION: FOR DIAGNOSTICS AND THERAPEUTICS
; FILE REFERENCE: 107196.132
; CURRENT APPLICATION NUMBER: US/09/248,796A
; CURRENT FILING DATE: 1999-02-12
; PRIOR APPLICATION NUMBER: US 60/074,725
; PRIOR FILING DATE: 1998-02-13
; PRIOR APPLICATION NUMBER: US 60/096,409
; PRIOR FILING DATE: 1998-08-13
; NUMBER OF SEQ ID NOS: 28208
; SEQ ID NO 4534
; LENGTH: 1635
; TYPE: DNA
; ORGANISM: Candida albicans
US-09-248-796A-4534

```
Alignment Scores:
Pred. No.:             3.33e-59      Length:        1635
Score:                 817.50        Matches:       188
Percent Similarity:    50.68%        Conservative:  108
Best Local Similarity: 32.19%        Mismatches:    181
Query Match:           17.74%        Indels:        107
DB:                    3             Gaps:          9
```

US-10-768-511-8 (1-901) x US-09-248-796A-4534 (1-1635)

```
Qy        308 SerSerIleIleProGluIleLysGlyAlaPhePheLysCysLeuValCysGlyHisSer 327
Db          1 ACCGGGGTTTTTCCACAATTGAAATATGTCAAATTTGATTGTTTGAAATGTGGA------ 54

Qy        328 ProProLeuValThrValValLysGlyArgValGluGluProThrArgCysGluLysPro 347
Db         55 GTGGTTTTGGTTCCATACGTTCAAGATTCTAATACTGAAGTGAAAATCTCATTCTGTACC 114

Qy        348 GluCysAlaAlaArgAsnAlaMetSerLeuIleHisAsnArgCysThrPheAlaAsnLys 367
Db        115 AATTGTCAAAGTAAAGGTCCTTTTAAATTGAATTCAGAGAAAACTTTATATAGAAACTAT 174

Qy        368 GlnIleValArgLeuGlnGluThrProAspAlaIleProGluGlyGluThrProHisThr 387
Db        175 CAAAGAATTACTTTACAAGAAGCTCCAGGTACTGTTCCAGCGGGTAGATTGCCAAGACAT 234

Qy        388 ValSerMetCysLeuTyrAsnThrMetValAspAlaValLysProGlyAspArgIleGlu 407
Db        235 CGTGAAGTCATATTATTATCTGATTTAGTTGATGTTGCTAAACCAGGTGAAGATATTGAA 294

Qy        408 ValThrGlyValPheLysAlaMetAlaValArgValGlyProAsnGlnArgThrLeuArg 427
Db        295 GTTACTGGTATTTACAAG-----------------AATAATTATGATGGGAATTTAAAT 336

Qy        428 Ala---------------LeuTyrLysThrTyrIleAspCysValHisValLysLysSer 442
Db        337 GCCAAGAATGGATTCCCGGTGTTTGCCACAATTTGGAAGCCAATTCCATTAGAAGA--- 393

Qy        443 AspArgGlyArgLeuGlnThrGluAspProMetGluMetAspLysGluAsnAspMetTyr 462
Db        394 -----------------------------------------AAAGAAAGTAGTGCATTT 411

Qy        463 AlaGlyTyrHisGluSerAspThrSerGluAlaAlaAsnGluAlaLysIleGlnLysLeu 482
Db        412 ATGGGT------GGTAATAATTTAGTCAATATGTGGACAGAAGAAGAAATACGTGAATTT 465

Qy        483 LysGluLeuSerLysLeuProGlyIleTyrAspArgLeuSerArgSerLeuAlaProSer 502
Db        466 AGAAAATTATCACACGAAAAAGGTATTATTGATAAAATCATTGCCTCAATGGCTCCATCC 525

Qy        503 IleTrpGluLeuGluAspIleLysLysGlyLeuLeuCysGlnLeuPheGlyGlyLysAla 522
Db        526 ATATATGGTCATAAAGATATTAAAACTGCCATTGCTTGTTCATTATTTGGTGGTGTGCCT 585

Qy        523 LysLysIleProSerGlyAlaSerPheArgGlyAspIleAsnValLeuLeuValGlyAsp 542
Db        586 AAAGATGTTAATGGGAAACTTTCCATTAGAGGTGATATTAATGTACTATTATTAGGGGAT 645

Qy        543 ProGlyThrSerLysSerGlnLeuLeuGlnTyrValHisLysIleAlaProArgGlyIle 562
Db        646 CCAGGTACTGCTAAATCGCAAATATTGAAATATGCTGAGAAAACTGCTAGTCGTGCTGTT 705

Qy        563 TyrThrSerGlyArgGlySerSerAlaValGlyLeuThrAlaTyrValThrLysAspPro 582
Db        706 TTTGCAACTGGTCAAGGTGCCTCGGCTGTTGGTCTTACAGCATCGGTTAGAAAAGACCCA 765

Qy        583 GluThrArgGluThrValLeuGluSerGlyAlaLeuValLeuSerAspArgGlyIleCys 602
Db        766 ATCACTCGTGAATGGACATTACAAGGTGGGGCTTTGGTGCTTGCAGATAAAGGGACTTGT 825

Qy        603 CysIleAspGluPheAspLysMetSerAspAsnAlaArgSerMetLeuHisGluValMet 622
Db        826 CTTATTGATGAATTTGATAAGATGAATGATCAAGATAGAACTTCTATTCATGAAGCTATG 885

Qy        623 GluGlnGlnThrValSerValAlaLysGlyGlyIleIleAlaSerLeuAsnAlaArgThr 642
```

```
Db        886 GAACAACAAAGTATATCTATTTCTAAAGCTGGTATCGTCACTACCTTACATGCCAGATGT 945

Qy        643 SerValLeuAlaCysAlaAsnProSerGlySerArgTyrAsnAlaArgLeuSerValIle 662
              :::|||:::|||  |||||||||:::|||   |||||||||:::   |||   :::
Db        946 GCTGTTATCGCTGCTGCCAATCCTAATGGAGGTCGATACAACTCTACATTACCATTACCA 1005

Qy        663 AspAsnIleGlnLeuProProThrLeuLeuSerArgPheAspLeuIleTyrLeuMetLeu 682
              :::|||:::   |||         :::||||||||||||||||:::::: ::::::
Db       1006 GAAAATGTCGATTTAACTTACCCAATTTTATCCAGATTTGATATTATGTGTATCGTGCGA 1065

Qy        683 AspLysProAspGluGlnAsnAspArgArgLeuAlaArgHisLeuValAlaLeuHisTyr 702
              |||      :::   ::::::|||  |||||||||      ::::::     |||
Db       1066 GATTTGGTCAATCCTGAATCTGATGAACGATTAGCTAGTTTTGTCATTGATTCCCATATG 1125

Qy        703 GluAsnTyrGluValSerLysGlnAspAlaLeuAspLeu--------------------- 715
                 ::::::      :::::::::|||  |||:::
Db       1126 AGATCTCATCCCACCAATGAAGAGGATATTTTAAATGCTAATGGTAAGGGTGGCTCTGAC 1185

Qy        715 ------------------------------------------------------------ 715

Db       1186 GCTCAAGACGATGATGAAAATATGGAAGACGAAGATGGTGTAGATCAACCATCAGCGGCA 1245

Qy        715 ------------------------------------------------------------ 715

Db       1246 AGAACTAGATCAGAAAGAATTGAACAATTAAATAAACAAAAAGAACAAGAAATATCACCT 1305

Qy        716 ---------GlnThrLeuThrAlaTyrIleThrTyrAlaArgGlnHisValHisProThr 732
                             |||   ||||||  |||||||||      :::   |||
Db       1306 ATTCCACAAGATTTATTAATTAAATATATTCAATATGCTCGAGTTAAAATTCAACCAAAA 1365

Qy        733 LeuSerAspGluAlaAlaGluAspLeuIleAsnGlyTyrValGluMetArgGlnLys--- 751
              |||                :::   |||         |||   ::::::|||::::::
Db       1366 TTACATCAAATGAATATGGATAAACTTGCTCGAGTATATGCTGATTTAAGAAAAGAAGCC 1425

Qy        752 ---------GlyAsnPheProGlySerSerLysLysValIleThrAlaThrProArgGln 768
                       |||::::::|||                        |||   |||
Db       1426 ATCACTACTGGTTCTTATCCA------------------------ATTACTGTTCGTCAT 1461

Qy        769 LeuGluSerMetIleArgIleSerGluAlaLeuAlaArgMetArgPheSerGluValVal 788
              ||||||||::::::|||||||:::|||:::   |||:::||||||  ||||||   |||
Db       1462 TTAGAAAGTATTTTAAGAATTGCTGAATCCTTTGCCAAAATGAGATTATCAGAATTTGTT 1521

Qy        789 GluLysValAspAlaAlaGluAlaValArgLeuLeuAspValAlaLeuGlnGlnSerAla 808
                 :::   |||         |||:::::::::
Db       1522 AGTCAAAATGATTTAAATCGAGCTATTAAAGTT--------------------------- 1554

Qy        809 ThrAspHisAlaThrGlyThrIleAspMetAspLeuIleThrThrGlyValSerAlaSer 828
                                 :::||||||      :::         |||:::
Db       1555 -----------------AGTATTGATAGTTTTATTGGTGCTCAAAAAGTGACAGTAAAA 1596

Qy        829 GluArgIleArg 832
              :::::::::|||
Db       1597 CAAAAACTTAGA 1608
```

RESULT 12
US-09-248-796A-5904
; Sequence 5904, Application US/09248796A
; Patent No. 6747137
; GENERAL INFORMATION:
; APPLICANT: Keith Weinstock et al
; TITLE OF INVENTION: NUCLEIC ACID AND AMINO ACID SEQUENCES RELATING TO CANDIDA ALBICAN
; TITLE OF INVENTION: FOR DIAGNOSTICS AND THERAPEUTICS
; FILE REFERENCE: 107196.132
; CURRENT APPLICATION NUMBER: US/09/248,796A
; CURRENT FILING DATE: 1999-02-12
; PRIOR APPLICATION NUMBER: US 60/074,725
; PRIOR FILING DATE: 1998-02-13
; PRIOR APPLICATION NUMBER: US 60/096,409
; PRIOR FILING DATE: 1998-08-13
; NUMBER OF SEQ ID NOS: 28208
; SEQ ID NO 5904
;   LENGTH: 1488
;   TYPE: DNA
;   ORGANISM: Candida albicans
;   FEATURE:
;   NAME/KEY: unsure
;   LOCATION: (2)
;   OTHER INFORMATION: Identity of nucleotide sequences at the above locations are unknc
US-09-248-796A-5904

Alignment Scores:

Pred. No.:	2.98e-52	Length:	1488
Score:	734.50	Matches:	184
Percent Similarity:	54.95%	Conservative:	88
Best Local Similarity:	37.17%	Mismatches:	170
Query Match:	15.94%	Indels:	53
DB:	3	Gaps:	13

US-10-768-511-8 (1-901) x US-09-248-796A-5904 (1-1488)

```
Qy        408 ValThrGlyValPheLysAlaMetAlaValArgValGlyProAsnGlnArgThrLeuArg 427
              ||||||||:::   |||::::::   |||||
Db         31 GTTACTGGTTTA---AAGTCTCTTGGTGTCAGA-----------------------GAT 63

Qy        428 AlaLeuTyrLysThrTyrIleAspCysValHisValLysLysSerAspArgGlyArgLeu 447
                    ||||||         ||||||      |||               :::
Db         64 TTGACATATAAGCTCGCCTTTGGAGCTTGCCACGTT---GCTTCCATGGTTAATAAGGCA 120

Qy        448 GlnThrGluAspProMetGluMetAsp---LysGluAsnAspMetTyrAlaGlyTyrHis 466
                 :::   :::|||:::|||      :::   :::::::::
Db        121 GGTGGGAATGAACAGCTTGAAGTTGATTTGAATGACCAAGAAGTTTTTCTT--------- 171

Qy        467 GluSerAspThrSerGluAlaAlaAsnGluAlaLysIleGlnLysLeuLysGluLeuSer 486
                             :::   ::::::|||::::::   :::|||||||||:::
Db        172 ---------------ACATCTTTGAGTGATGCTGAAGTATTACAATTAAAGGAAATGGTC 216

Qy        487 LysLeuProGlyIleTyrAspArgLeuSerArgSerLeuAlaProSerIleTrpGluLeu 506
              |||         |||||||||:::|||         |||:::||||||:::::::::
Db        217 AAGGATGAACATATTTATGACAAGCTTGTGAACTCAATTGCACCAGCTGTGTTTGGCCAC 276

Qy        507 GluAspIleLysLysGlyLeuLeuCysGlnLeuPheGlyGlyLysAlaLysLysIlePro 526
              |||   |||||||||||:::|||   ||||||   |||||        |||:::
Db        277 GAAGTGATCAAAAAGGGAATTCTTTTGCAATTATTAGGCGGTGTTCACAAACAAACAGTA 336

Qy        527 SerGlyAlaSerPheArgGlyAspIleAsnValLeuLeuValGlyAspProGlyThrSer 546
                 |||   :::   ||||||||||||:::   :::|||||||||||   ||||||
Db        337 GATGGAATCAACTTAAGAGGGGATATAAACATTTGTATTGTTGGTGACCCATCCACTTCG 396

Qy        547 LysSerGlnLeuLeuGlnTyrValHisLysIleAlaProArgGlyIleTyrThrSerGly 566
              ||||||||   |||:::||||||         :::||||||   :::||||||||||||
Db        397 AAATCACAGTTTTTGAAGTATGTCTGTGGGTTTTCGCCTCGTGCTGTCTACACTTCTGGT 456

Qy        567 ArgGlySerSerAlaValGlyLeuThrAlaTyrValThrLysAspProGluThrArgGlu 586
              :::   |||||||||   ||||||||||||   |||   ||||||   |||:::   |||
Db        457 AAGGCTTCTTCTGCTGCTGGGTTGACAGCGGCTGTAGTGAAGGACGAAGAAAGCGGAGAA 516

Qy        587 ThrValLeuGluSerGlyAlaLeuValLeuSerAspArgGlyIleCysCysIleAspGlu 606
                 :::|||:::|||||||||:::|||:::|||   |||||||||   |||||||||
Db        517 TACACTATTGAAGCAGGTGCTTTGATGTTGGCAGATAACGGTATTTGTGCAATTGATGAA 576

Qy        607 PheAspLysMetSerAspAsnAlaArgSerMetLeuHisGluValMetGluGlnGlnThr 626
              ||||||||||||         :::      :::||||||   |||||||||||||||
Db        577 TTTGACAAGATGGATATTGCTGACCAAGTTGCTATTCACGAAGCCATGGAACAACAAACA 636

Qy        627 ValSerValAlaLysGlyGlyIleIleAlaSerLeuAsnAlaArgThrSerValLeuAla 646
              :::|||:::||||||   ||||||   |||:::||||||||||||||||:::||||||
Db        637 ATCTCCATTGCAAAAGCTGGTATTCATGCTACCTTAAATGCCCGTACATCTATCTTAGCA 696

Qy        647 CysAlaAsnProSerGlySerArgTyrAsnAlaArgLeuSerValIleAspAsnIleGln 666
                 |||||||||   |||   |||||||||   :::|||   :::      |||:::
Db        697 GCTGCTAACCCTATTGGGGGTAGATATAATAGAAAATTGGGTCTTCGTTCTAATTTGAAT 756

Qy        667 LeuProProThrLeuLeuSerArgPheAspLeuIleTyrLeuMetLeuAspLysProAsp 686
              :::         ::::::|||||||||||||||   :::::::::||||||      :::
Db        757 ATGACAGCACCAATTATGTCGAGATTTGATTTGTTCTTTGTTGTTCTTGATGATTGTAAT 816
```

```
Qy       687 GluGlnAsnAspArgArgLeuAlaArgHisLeuValAlaLeuHisTyrGluAsnTyrGlu 706
Db       817 GAAAGAATTGATACTCAATTAGCATCACACATTGTTGATTTGCAC--------------- 861

Qy       707 ValSerLysGlnAspAlaLeuAsp------------LeuGlnThrLeuThrAlaTyrIle 722
Db       862 ATGCTTAGAGACGATGCAATTGATCCACCTTACTCAGCCGAACAATTAGCAAGATACATT 921

Qy       723 ThrTyrAlaArgGlnHisValHisProThrLeuSerAspGluAlaAlaGluAspLeuIle 742
Db       922 AAATACGCAAAGACT---TTCAAGCCTAGAATGACTAAAGAAGCTAGAGACTTTTTGGTC 978

Qy       743 AsnGlyTyrValGluMetArgGlnLysGlyAsnPheProGlySerSerLysLysValIle 762
Db       979 ACTAGATACAAGGAATTGAGAGAA---GATGATGCTCAAGGTTTAGGGAGGTCCTCGTAC 1035

Qy       763 ThrAlaThrProArgGlnLeuGluSerMetIleArgIleSerGluAlaLeuAlaArgMet 782
Db      1036 AGAATTACTGTTAGACAATTGGAGTCCATGATTAGATTGTCAGAAGCTATAGCCAGAGCC 1095

Qy       783 ArgPheSerGluValValGluLysValAspAlaAlaGluAlaValArgLeuLeuAspVal 802
Db      1096 AATTGTACCGAAGAAATAACGCCAAGTTTTGTTGCTGAGGCCTATGATTTGTTGAAGCAA 1155

Qy       803 Ala------------------------LeuGlnGlnSerAlaThrAspHisAlaThrGly 814
Db      1156 TCTATTATCAGAGTCGAAATGGATGACATCGAAATGGATGATGAAGAAAGACCAACTGGA 1215

Qy       815 ------------------------ThrIleAspMetAspLeuIleThrThrGlyValSer 826
Db      1216 GAACCGGGGCAAGCACAAGAACAAGATCAGCCACAGGACCAAGATACTCAGCAATCGCAA 1275

Qy       827 AlaSerGluArgIleArgArgAlaAsnLeu------------LeuAlaAlaLeuArgGlu 842
Db      1276 CCTTCTGAAAGAGTACCTCCAGTGTCAATTAGTTATGACAAATATGTTTCTATAATGAAT 1335

Qy       843 LeuIleAlaAspLysIleSer---------ProGlySerSerSerGlyLeuLysThrSer 859
Db      1336 ATGTTAGTCAAGAAAATAACCGAGGACGAAAAAAATGGAGGTGATGGACTTTCGGCTGAT 1395

Qy       860 GlnLeuLeuGluAspIleArgSerGlnSerSerValAspValSer 874
Db      1396 ACTTTAGTTGAATGGTATTTGCTTCAAAAGGAAGACGAGATTTCT 1440
```

RESULT 13
US-09-248-796A-4570
; Sequence 4570, Application US/09248796A
; Patent No. 6747137
; GENERAL INFORMATION:
; APPLICANT: Keith Weinstock et al
; TITLE OF INVENTION: NUCLEIC ACID AND AMINO ACID SEQUENCES RELATING TO CANDIDA ALBICAN
; TITLE OF INVENTION: FOR DIAGNOSTICS AND THERAPEUTICS
; FILE REFERENCE: 107196.132
; CURRENT APPLICATION NUMBER: US/09/248,796A
; CURRENT FILING DATE: 1999-02-12
; PRIOR APPLICATION NUMBER: US 60/074,725
; PRIOR FILING DATE: 1998-02-13
; PRIOR APPLICATION NUMBER: US 60/096,409
; PRIOR FILING DATE: 1998-08-13
; NUMBER OF SEQ ID NOS: 28208
; SEQ ID NO 4570
;   LENGTH: 996
;   TYPE: DNA
;   ORGANISM: Candida albicans
US-09-248-796A-4570

Alignment Scores:

Pred. No.:	4.42e-47	Length:	996
Score:	670.00	Matches:	155
Percent Similarity:	58.36%	Conservative:	58
Best Local Similarity:	42.47%	Mismatches:	107
Query Match:	14.54%	Indels:	45
DB:	3	Gaps:	7

US-10-768-511-8 (1-901) x US-09-248-796A-4570 (1-996)

```
Qy       318 PhePheLysCysLeuValCysGlyHisSerProProLeuValThrValValLysGlyArg 337
Db        16 TTTTCCAAGAAATCAATCACAAAACATTCACTCCCATTA--------------------- 54

Qy       338 ValGluGluProThrArgCysGluLysProGluCysAlaAlaArgAsnAlaMetSerLeu 357
Db        55 ------------ACCGAATGTAATTCACCATCAGGTGTTAATGATAACAATAAGGGTCAA 102

Qy       358 IleHisAsnArgCysThrPheAlaAsnLys------------GlnIleValArgLeuGln 373
Db       103 TTATTC---ATGTCCACCCGTGCATCAAAGTTATCTGCTTTCCAAGAAGTTAAAATCCAA 159

Qy       374 GluThrProAspAlaIleProGluGlyGluThrProHisThrValSerMetCysLeuTyr 393
Db       160 GAATTATCAAGTCAAGTCCCTGTTGGACATATTCCAAGAAGTTTAACTGTCCATGTCAAT 219

Qy       394 AsnThrMetValAspAlaValLysProGlyAspArgIleGluValThrGlyValPheLys 413
Db       220 GGTGATTTGGTCAGATCAATGAATCCTGGTGATACCGTTGATTTATCTGGAATATTCATG 279

Qy       414 AlaMetAlaValArgValGlyProAsnGlnArgThrLeuArgAla------LeuTyrLys 431
Db       280 CCAAGTCCATAC---------ACTGGATATCGTGCATTGAAAGCTGGGTTGTTGACTGAA 330

Qy       432 ThrTyrIleAspCysValHisValLysLysSerAspArgGlyArgLeuGlnThrGluAsp 451
Db       331 ACCTATTTAGAAGCTCAGCATGTTAAACAGCACAAAAAG--------------------- 369

Qy       452 ProMetGluMetAspLysGluAsnAspMetTyrAlaGlyTyrHisGluSerAspThrSer 471
Db       370 ---------------------------------------------CAATATGATTCAATG 384

Qy       472 GluAlaAlaAsnGluAlaLysIleGlnLysLeuLysGluLeuSerLysLeuProGlyIle 491
Db       385 ACATTATCGAGTCAAGCACAA---GACAAAATCGACGAGCTTTTACTTCAAGGTGATGTT 441

Qy       492 TyrAspArgLeuSerArgSerLeuAlaProSerIleTrpGluLeuGluAspIleLysLys 511
Db       442 TACAATAAGTTGGCCAAATCCATTGCTCCGGAAATTTATGGTCACTTGGATGTTAAGAAA 501

Qy       512 GlyLeuLeuCysGlnLeuPheGlyGlyLysAlaLysLysIleProSerGlyAlaSerPhe 531
Db       502 ATTTTATTATTGTTATTATGTGGTGGTGTCACCAAAGAAATCGGTGATGGGTTGAAAATT 561

Qy       532 ArgGlyAspIleAsnValLeuLeuValGlyAspProGlyThrSerLysSerGlnLeuLeu 551
Db       562 AGAGGTGATATTAATGTTTGTCTTATGGGTGATCCGGGGGTTGCTAAATCACAATTATTG 621

Qy       552 GlnTyrValHisLysIleAlaProArgGlyIleTyrThrSerGlyArgGlySerSerAla 571
Db       622 AAAGCCATTGGTAAGATTGCTCCAAGATCGGTCTATACTACTGGTAGAGGTTCTTCTGGT 681

Qy       572 ValGlyLeuThrAlaTyrValThrLysAspProGluThrArgGluThrValLeuGluSer 591
Db       682 GTAGGGTTGACGGCAGCTGTTATGAGAGATCCAATTACTGATGAAATGGTGTTGGAAGGT 741

Qy       592 GlyAlaLeuValLeuSerAspArgGlyIleCysCysIleAspGluPheAspLysMetSer 611
Db       742 GGGGCTTTGGTGTTGGCCGATAACGGTATTTGTTGTATTGATGAATTTGATAAAATGGAC 801

Qy       612 AspAsnAlaArgSerMetLeuHisGluValMetGluGlnGlnThrValSerValAlaLys 631
Db       802 GAAAGCGACAGAACGGCGATTCATGAAGTTATGGAACAACAAACCATCTCTATTGCCAAA 861

Qy       632 GlyGlyIleIleAlaSerLeuAsnAlaArgThrSerValLeuAlaCysAlaAsnProSer 651
Db       862 GCTGGTATCACTACTACTTTGAATGCCAGAACTTCGATCTTGGCAGCAGCCAATCCATTA 921

Qy       652 GlySerArgTyrAsnAlaArgLeuSerValIleAspAsnIleGlnLeuPro-ProThrLe 671
```

```
Db        922 TATGGTAGATATAATCCAAGATTGTCTCCACATGAAAATATCAATTTACCAGCCAGCTTT 981

Qy        671 uLeuSerArgPhe 675
              |   |||||||||
Db        982 GTGGTCCCGTTTT 994

RESULT 14
US-08-916-421B-1/c
; Sequence 1, Application US/08916421B
; Patent No. 6503729
; GENERAL INFORMATION:
;  APPLICANT: Bult et al.
;  TITLE OF INVENTION: Complete Genome Sequence of the Methanogenic Archaeon, Methanococ
; Patent No. 6503729
;  TITLE OF INVENTION:  jannaschii
;  FILE REFERENCE: PB275
;  CURRENT APPLICATION NUMBER: US/08/916,421B
;  CURRENT FILING DATE:  1997-08-22
;  PRIOR APPLICATION NUMBER: US 60/024,428
;  PRIOR FILING DATE: 1996-08-22
;  NUMBER OF SEQ ID NOS: 3
;  SOFTWARE: PatentIn version 3.1
; SEQ ID NO 1
;   LENGTH: 1664976
;   TYPE: DNA
;   ORGANISM: Methanococcus jannaschii
;   FEATURE:
;   NAME/KEY: misc_feature
;   LOCATION: (28222)..(28222)
;   OTHER INFORMATION: n equals a, t, c, or g
;   NAME/KEY: misc_feature
;   LOCATION: (28257)..(28258)
;   OTHER INFORMATION: n equals a, t, c, or g
;   NAME/KEY: misc_feature
;   LOCATION: (84773)..(84773)
;   OTHER INFORMATION: n equals a, t, c, or g
;   NAME/KEY: misc_feature
;   LOCATION: (84808)..(84808)
;   OTHER INFORMATION: n equals a, t, c, or g
;   NAME/KEY: misc_feature
;   LOCATION: (84812)..(84812)
;   OTHER INFORMATION: n equals a, t, c, or g
;   NAME/KEY: misc_feature
;   LOCATION: (98120)..(98120)
;   OTHER INFORMATION: n equals a, t, c, or g
;   NAME/KEY: misc_feature
;   LOCATION: (98159)..(98159)
;   OTHER INFORMATION: n equals a, t, c, or g
;   NAME/KEY: misc_feature
;   LOCATION: (98239)..(98239)
;   OTHER INFORMATION: n equals a, t, c, or g
;   NAME/KEY: misc_feature
;   LOCATION: (98266)..(98266)
;   OTHER INFORMATION: n equals a, t, c, or g
;   NAME/KEY: misc_feature
;   LOCATION: (98343)..(98343)
;   OTHER INFORMATION: n equals a, t, c, or g
;   NAME/KEY: misc_feature
;   LOCATION: (103998)..(103998)
;   OTHER INFORMATION: n equals a, t, c, or g
;   NAME/KEY: misc_feature
;   LOCATION: (148948)..(148948)
;   OTHER INFORMATION: n equals a, t, c, or g
;   NAME/KEY: misc_feature
;   LOCATION: (163385)..(163385)
;   OTHER INFORMATION: n equals a, t, c, or g
;   NAME/KEY: misc_feature
;   LOCATION: (191989)..(191989)
;   OTHER INFORMATION: n equals a, t, c, or g
;   NAME/KEY: misc_feature
;   LOCATION: (191995)..(191995)
;   OTHER INFORMATION: n equals a, t, c, or g
;   NAME/KEY: misc_feature
;   LOCATION: (231980)..(231980)
;   OTHER INFORMATION: n equals a, t, c, or g
;   NAME/KEY: misc_feature
;   LOCATION: (234187)..(234187)
;   OTHER INFORMATION: n equals a, t, c, or g
;   NAME/KEY: misc_feature
;   LOCATION: (234220)..(234220)
;   OTHER INFORMATION: n equals a, t, c, or g
;   NAME/KEY: misc_feature
;   LOCATION: (234814)..(234814)
;   OTHER INFORMATION: n equals a, t, c, or g
;   NAME/KEY: misc_feature
;   LOCATION: (309398)..(309398)
;   OTHER INFORMATION: n equals a, t, c, or g
;   NAME/KEY: misc_feature
;   LOCATION: (309418)..(309418)
;   OTHER INFORMATION: n equals a, t, c, or g
;   NAME/KEY: misc_feature
;   LOCATION: (312837)..(312837)
;   OTHER INFORMATION: n equals a, t, c, or g
;   NAME/KEY: misc_feature
;   LOCATION: (312993)..(312993)
;   OTHER INFORMATION: n equals a, t, c, or g
;   NAME/KEY: misc_feature
;   LOCATION: (319226)..(319226)
;   OTHER INFORMATION: n equals a, t, c, or g
;   NAME/KEY: misc_feature
;   LOCATION: (559167)..(559167)
;   OTHER INFORMATION: n equals a, t, c, or g
;   NAME/KEY: misc_feature
;   LOCATION: (559241)..(559241)
;   OTHER INFORMATION: n equals a, t, c, or g
;   NAME/KEY: misc_feature
;   LOCATION: (600992)..(600992)
;   OTHER INFORMATION: n equals a, t, c, or g
;   NAME/KEY: misc_feature
;   LOCATION: (622708)..(622708)
;   OTHER INFORMATION: n equals a, t, c, or g
;   NAME/KEY: misc_feature
;   LOCATION: (657081)..(657081)
;   OTHER INFORMATION: n equals a, t, c, or g
;   NAME/KEY: misc_feature
;   LOCATION: (657203)..(657203)
;   OTHER INFORMATION: n equals a, t, c, or g
;   NAME/KEY: misc_feature
;   LOCATION: (674435)..(674435)
;   OTHER INFORMATION: n equals a, t, c, or g
;   NAME/KEY: misc_feature
;   LOCATION: (682442)..(682442)
;   OTHER INFORMATION: n equals a, t, c, or g
;   NAME/KEY: misc_feature
;   LOCATION: (713652)..(713652)
;   OTHER INFORMATION: n equals a, t, c, or g
;   NAME/KEY: misc_feature
;   LOCATION: (741684)..(741684)
;   OTHER INFORMATION: n equals a, t, c, or g
;   NAME/KEY: misc_feature
;   LOCATION: (779455)..(779455)
;   OTHER INFORMATION: n equals a, t, c, or g
;   NAME/KEY: misc_feature
;   LOCATION: (779676)..(779676)
;   OTHER INFORMATION: n equals a, t, c, or g
;   NAME/KEY: misc_feature
;   LOCATION: (855539)..(855539)
;   OTHER INFORMATION: n equals a, t, c, or g
;   NAME/KEY: misc_feature
;   LOCATION: (871619)..(871619)
;   OTHER INFORMATION: n equals a, t, c, or g
;   NAME/KEY: misc_feature
;   LOCATION: (1084830)..(1084830)
;   OTHER INFORMATION: n equals a, t, c, or g
;   NAME/KEY: misc_feature
;   LOCATION: (1096846)..(1096846)
```

```
;   OTHER INFORMATION: n equals a, t, c, or g
;   NAME/KEY: misc_feature
;   LOCATION: (1119881)..(1119881)
;   OTHER INFORMATION: n equals a, t, c, or g
;   NAME/KEY: misc_feature
;   LOCATION: (1130881)..(1130881)
;   OTHER INFORMATION: n equals a, t, c, or g
;   NAME/KEY: misc_feature
;   LOCATION: (1310988)..(1310988)
;   OTHER INFORMATION: n equals a, t, c, or g
;   NAME/KEY: misc_feature
;   LOCATION: (1313224)..(1313224)
;   OTHER INFORMATION: n equals a, t, c, or g
;   NAME/KEY: misc_feature
;   LOCATION: (1349473)..(1349473)
;   OTHER INFORMATION: n equals a, t, c, or g
;   NAME/KEY: misc_feature
;   LOCATION: (1349491)..(1349491)
;   OTHER INFORMATION: n equals a, t, c, or g
;   NAME/KEY: misc_feature
;   LOCATION: (1470091)..(1470091)
;   OTHER INFORMATION: n equals a, t, c, or g
;   NAME/KEY: misc_feature
;   LOCATION: (1569020)..(1569020)
;   OTHER INFORMATION: n equals a, t, c, or g
;   NAME/KEY: misc_feature
;   LOCATION: (1602912)..(1602912)
;   OTHER INFORMATION: n equals a, t, c, or g
;   NAME/KEY: misc_feature
;   LOCATION: (1603734)..(1603734)
;   OTHER INFORMATION: n equals a, t, c, or g
;   NAME/KEY: misc_feature
;   LOCATION: (1637998)..(1637998)
;   OTHER INFORMATION: n equals a, t, c, or g
;   NAME/KEY: misc_feature
;   LOCATION: (1664854)..(1664855)
;   OTHER INFORMATION: n equals a, t, c, or g
US-08-916-421B-1

Alignment Scores:
Pred. No.:              1.19e-40     Length:        1664976
Score:                  656.00       Matches:       202
Percent Similarity:     46.75%       Conservative:  143
Best Local Similarity:  27.37%       Mismatches:    258
Query Match:            14.23%       Indels:        135
DB:                     3            Gaps:          26

US-10-768-511-8 (1-901) x US-08-916-421B-1 (1-1664976)

Qy      177 TyrValTrpGly-ThrAsnValAsnIleProAsp---------ValLeuArgAlaIleAr 193
Db    10758 TTCGTATGGGGATGTAGATATGGAATTGAGAGATGAAGACTTAATTTTAGAGGAAGTTAG 10699

Qy      193 gArgPheLeuHisAsnTyrArgSerSerAlaHis------------------------- 204
Db    10698 GGATTATTTAACTGCCTATTTAAGAAATATTCATCAGGAAGATATTATCTTAGACAATGA 10639

Qy      205 -------------AspLeuAsnSerLysTyr--------------------IleGlnIl 213
Db    10638 AAGAGTTGTAGTTGATTTAAATCAACTCTACAATTACGGATTAATGGAATTTGTAGAATT 10579

Qy      213 eIleGluGluThrValGluArgGluGluAspThrLeuAsnIleAspMetSerAspIleTy 233
Db    10578 TTTAATAAATAATCCACAAAAAGGGATTGATTTTATAAAAGAGTGCTACAACGATGCTTA 10519

Qy      233 rAspHisAspProAspLeuTyrAlaLysIleValArgTyrProLeuAspIleIleProLe 253
Db    10518 T------------------TACACCTTAAGAAATGAATATCCAACAAACGTAATA----- 10482

Qy      253 uLeuAspThrGluCysGlnGluValAlaThrSerLeuLeuPro---ThrPheGluLysHi 272
Db    10481 ---------------------ATTGCTGTAAAAAATCTCCCAAAAATTTTTAAAACTAC 10444

Qy      272 sIleGluAlaArgProPheAsnLeuLysAlaSerValHisMetArgGluLeuAsnProSe 292
Db    10443 AAGAAAAGGGAAAATTTTCACAATAGAG-----------------GATATTAGGAGCAA 10402

Qy      292 rAspIleAspLysLeuValSerValLysGlyMetValIleArgCysSerSerIleIlePr 312
Db    10401 AACTTTAGGAAAATTAGTTGAATTTGAAGGAATCATTGTTATGGCATCAAAAATAAGACC 10342

Qy      312 oGluIleLysGlyAlaPhePheLysCysLeuValCysGlyHisSerProProLeuValTh 332
Db    10341 AATGCTTAAAAAAGCTTATTATATATGTCCAAAATGTGGAAGAGAA------------- 10296

Qy      332 rValValLysGlyArgValGluGluProThrArgCysGluLysProGluCysAlaAlaAr 352
Db    10295 -GTTGTTAGAGAAATAGATATCTTAAACACTGACTCTGAAAAAGCTGTTTGTGAATGTGG 10237

Qy      352 gAsnAlaMetSerLeuIleHisAsnArgCysThrPheAlaAsnLysGlnIleValArgLe 372
Db    10236 AGCTGAGTTAAACCTAATTGAGCATAAATCAATATATACTGACTTCCAAGAGATTAAAGT 10177

Qy      372 uGlnGluThrProAspAlaIle------ProGluGlyGluThrProHisThrValSerMe 390
Db    10176 TCAACAGCCATTGGATTTAATGGAAAATCCTGAA---GAGCCGCCAAAATACATAACTGT 10120

Qy      390 tCysLeuTyrAsnThrMetValAspAlaValLysProGly---------AspArgIleGl 407
Db    10119 GTTCTTAGAAAACAGT-----------------CCTGGAATATATGCTGGAAGGGTGAA 10078

Qy      407 uValThrGlyValPheLysAlaMetAlaValArgValGlyProAsnGlnArgThrLeuAr 427
Db    10077 GATAACTGGCATC------------CCAATAAAAGTTAAGAAGAGTAAAAAGCTA----- 10035

Qy      427 gAlaLeuTyrLysThrTyrIleAspCysValHisValLysLysSerAspArgGlyArgLe 447
Db    10034 -CCAATCTATGACATACATGTTAAAGCCCTACACTGTGAGGTTTTGGAT----------- 9987

Qy      447 uGlnThrGluAspProMetGluMetAspLysGluAsnAspMetTyrAlaGlyTyrHisGl 467
Db     9987 ------------------------------------------------------------ 9987

Qy      467 uSerAspThrSerGluAlaAlaAsnGluAlaLysIleGlnLysLeuLysGluLeuSerLy 487
Db     9986 -GGGGAAGTAAAAATTAAATTAACAAATTCAGATATTGAAAATATTAAAAAAATAGCTAA 9928

Qy      487 sLeuProGlyIleTyrAspArgLeuSerArgSerLeuAlaProSerIleTrpGluLeuGl 507
Db     9927 AAGAAAAGATGTTGTTAATATACTTGCAGATAGATTAATTCCAGAGATTAAGGGGCATTC 9868

Qy      507 uAspIleLysLysGlyLeuLeuCysGlnLeuPheGlyGlyLysAlaLysLysIleProSe 527
Db     9867 TGCAATAAAGAAGGCTGTCTTACTACAACAGATAAAGGGAGTTAAAAAA----------- 9819

Qy      527 rGlyAlaSerPheArgGlyAspIleAsnValLeuLeuValGlyAspProGlyThrSerLy 547
Db     9818 ----CCTGGAAAGAGGGCTGATATTCATATATTATTAATCACAGACCCAGGAATTGGAAA 9763

Qy      547 sSerGlnLeuLeuGlnTyrValHisLysIleAlaProArgGlyIleTyrThrSerGlyAr 567
Db     9762 AACAGTCATACTAAGAAAGATTGCTGAAATC---CCAGGAAATTTATACGGTTCTGTTAC 9706

Qy      567 gGlySerSerAlaValGlyLeuThrAlaTyrValThrLysAsp---ProGluThrArgGl 586
Db     9705 TACCGCCACTGGTGTTGGGCTAACTGCCGCTGTTGTTAGGGAGAAGACAGAGATTGGTGA 9646

Qy      586 uThr------ValLeuGluSerGlyAlaLeuValLeuSerAspArgGlyIleCysCysIl 604
Db     9645 AGATACATGGGTTATTAAGCCAGGTTTGTTAGTTAAAGCCCACAAAGGAACTGCATGTAT 9586

Qy      604 eAspGluPheAspLysMetSerAspAsnAlaArgSerMetLeuHisGluValMetGluGl 624
Db     9585 AGATGAGCTA---ACTGTTAATAAAGAACTGCAGAGCTATGTTTTGGAGGCTATGGAGAG 9529

Qy      624 nGlnThrValSerValAlaLysGlyGlyIleIleAlaSerLeuAsnAlaArgThrSerVa 644
```

```
              ||||||:::   ::::::|||||||||||    |||    |||    |||       :::::
Db       9528 TCAGACAATCCATATTAGTAAAGGAGGAATAAATGCAAAACTTCCAGCTGAATGTGCAAT 9469

Qy        644 lLeuAlaCysAlaAsnProSerGlySerArgTyrAsnAlaArgLeuSerValIleAspAs 664
               :||||||      ||||||           |||:::|||      :::||||||    :::
Db       9468 TTTAGCCGCTTGCAATCCAAGATGGGGAAGATTCAACCCAGAGGTTTCTGTAGCTGAGCA 9409

Qy        664 nIleGlnLeuProProThrLeuLeuSerArgPheAspLeuIleTyrLeuMetLeuAspLy 684
              |||   :::|||      ||||||||||||||||||||||||:::   :::    |||
Db       9408 GATAAATATTCCAGCCCCATTATTGAGTAGATTTGACCTAATATTTCCAATTAGAGATGT 9349

Qy        684 sProAspGluGlnAsnAspArgArgLeuAlaArgHisLeuValAlaLeuHis------Ty 702
                 |||:::      |||:::   :::|||   ::::::|||   ||||||        ||
Db       9348 TTCTGATAAGGATAAGGATAAAGATATCGCTGAATACATTGTAGATTTGCATAGAGCATA 9289

Qy        702 rGluAsnTyrGluValSerLysGlnAspAlaLeuAspLeu-------------------- 715
              |   :::   :::::::::::::::            ||||||
Db       9288 TTTAGATGAGAAAATAAACAGGGAGATGGGTTTAGATTACCTTGAGGTTGATGGAGTTAA 9229

Qy        716 ----------GlnThrLeuThrAlaTyrIleThrTyrAlaArgGlnHisValHisProTh 732
                        :::   :::         ||||||   ||||||||:|||         |||
Db       9228 AATAGACAAGGAATTTATAATAAAATATATCTACTACGCAAGACAGAAA---AAGCCGAT 9172

Qy        732 rLeuSerAspGluAlaAlaGluAspLeuIleAsnGlyTyrValGluMetArgGlnLysGl 752
               :::|||:::::|||    |||      :::|||   ||||| |||||||||:::|||
Db       9171 AATTAGTGAAAAGGCTAAGGAATTATTTGTTAATTATTATGTAGAGATGAGAAAAAAACA 9112

Qy        752 yAsnPheProGlySerSerLysLysValIleThrAlaThrProArgGlnLeuGluSerMe 772
                                                   |||   ||||||||||||:::
Db       9111 TCAGATA-----------------------------ACTGCAAGGCAGTTAGAGGCTGC 9082

Qy        772 tIleArgIleSerGluAlaLeuAlaArgMetArgPheSerGluValValGluLysValAs 792
                |||||||||:::||||||      |||:::   :::         :::||||||::::::   ||
Db       9081 TATAAGGATTGCTGAAGCTCATGCAAAGGCAAAGTTAAAGGATGTTGTTGATGAGGAAGA 9022

Qy        792 pAlaAlaGluAlaValArgLeuLeuAspValAlaLeuGlnGlnSerAlaThrAspHisAl 812
              ||||    ||||||:::    ::::::            |||:::::::   |||    |||
Db       9021 TGCAAAGGAGGCAATAAATATAATAACTGAATGCTTAAAAGAAATTGCCTATGACCCAGA 8962

Qy        812 aThrGlyThrIleAspMetAspLeuIleThrThrGlyValSerAlaSerGluArgIleAr 832
                ||||||      |||:::|||   |||      |||||||||        ||||||    ::
Db       8961 GACTGGAATATTTGATGTTGATAAAATT---TTAGGGGTGTCTAAGAAAGAGAGAGACAA 8905

Qy        832 gArgAlaAsnLeuLeuAlaAlaLeuArgGluLeuIle----------------------- 844
              :          :::            ::::::||||||
Db       8904 ATTAACAACCGTCTATGAGATAATTAAGGAGTTATCTGAAAAATCAGAACTTGTTGAACA 8845

Qy        845 -AlaAspLysIleSerProGlySerSerSerGlyLeuLysThrSerGlnLeu 861
                    |||                    |||:::|||          :::|||
Db       8844 TGAAGATATAGCTGAAGAAGCAAAAAAGAAAGGTATTAAAGAGGATGAGTTG 8793
```

RESULT 15
US-09-692-570-1/c
; Sequence 1, Application US/09692570
; Patent No. 6797466
; GENERAL INFORMATION:
; APPLICANT: Bult et al.
; TITLE OF INVENTION: Complete Genome Sequence of the Methanogenic Archaeon, Methanococ
; Patent No. 6797466
; TITLE OF INVENTION: jannaschii
; FILE REFERENCE: PB275C1
; CURRENT APPLICATION NUMBER: US/09/692,570
; CURRENT FILING DATE: 2003-01-14
; PRIOR APPLICATION NUMBER: US 60/024,428
; PRIOR FILING DATE: 1996-08-22
; PRIOR APPLICATION NUMBER: US 08/916,421
; PRIOR FILING DATE: 1997-08-22
; NUMBER OF SEQ ID NOS: 20
; SOFTWARE: PatentIn version 3.1
; SEQ ID NO 1
;   LENGTH: 1664976
;   TYPE: DNA
;   ORGANISM: Methanococcus jannaschii

;   FEATURE:
;   NAME/KEY: misc_feature
;   LOCATION: (28222)..(28222)
;   OTHER INFORMATION: n equals a, t, c, or g
;   FEATURE:
;   NAME/KEY: misc_feature
;   LOCATION: (28257)..(28258)
;   OTHER INFORMATION: n equals a, t, c, or g
;   FEATURE:
;   NAME/KEY: misc_feature
;   LOCATION: (84773)..(84773)
;   OTHER INFORMATION: n equals a, t, c, or g
;   FEATURE:
;   NAME/KEY: misc_feature
;   LOCATION: (84808)..(84808)
;   OTHER INFORMATION: n equals a, t, c, or g
;   FEATURE:
;   NAME/KEY: misc_feature
;   LOCATION: (84812)..(84812)
;   OTHER INFORMATION: n equals a, t, c, or g
;   FEATURE:
;   NAME/KEY: misc_feature
;   LOCATION: (98120)..(98120)
;   OTHER INFORMATION: n equals a, t, c, or g
;   FEATURE:
;   NAME/KEY: misc_feature
;   LOCATION: (98159)..(98159)
;   OTHER INFORMATION: n equals a, t, c, or g
;   FEATURE:
;   NAME/KEY: misc_feature
;   LOCATION: (98239)..(98239)
;   OTHER INFORMATION: n equals a, t, c, or g
;   FEATURE:
;   NAME/KEY: misc_feature
;   LOCATION: (98266)..(98266)
;   OTHER INFORMATION: n equals a, t, c, or g
;   FEATURE:
;   NAME/KEY: misc_feature
;   LOCATION: (98343)..(98343)
;   OTHER INFORMATION: n equals a, t, c, or g
;   FEATURE:
;   NAME/KEY: misc_feature
;   LOCATION: (103998)..(103998)
;   OTHER INFORMATION: n equals a, t, c, or g
;   FEATURE:
;   NAME/KEY: misc_feature
;   LOCATION: (148948)..(148948)
;   OTHER INFORMATION: n equals a, t, c, or g
;   FEATURE:
;   NAME/KEY: misc_feature
;   LOCATION: (163385)..(163385)
;   OTHER INFORMATION: n equals a, t, c, or g
;   FEATURE:
;   NAME/KEY: misc_feature
;   LOCATION: (191989)..(191989)
;   OTHER INFORMATION: n equals a, t, c, or g
;   FEATURE:
;   NAME/KEY: misc_feature
;   LOCATION: (191995)..(191995)
;   OTHER INFORMATION: n equals a, t, c, or g
;   FEATURE:
;   NAME/KEY: misc_feature
;   LOCATION: (231980)..(231980)
;   OTHER INFORMATION: n equals a, t, c, or g
;   FEATURE:
;   NAME/KEY: misc_feature
;   LOCATION: (234187)..(234187)
;   OTHER INFORMATION: n equals a, t, c, or g
;   FEATURE:
;   NAME/KEY: misc_feature
;   LOCATION: (234220)..(234220)
;   OTHER INFORMATION: n equals a, t, c, or g
;   FEATURE:

```
;   NAME/KEY: misc_feature
;   LOCATION: (234814)..(234814)
;   OTHER INFORMATION: n equals a, t, c, or g
;   FEATURE:
;   NAME/KEY: misc_feature
;   LOCATION: (309398)..(309398)
;   OTHER INFORMATION: n equals a, t, c, or g
;   FEATURE:
;   NAME/KEY: misc_feature
;   LOCATION: (309418)..(309418)
;   OTHER INFORMATION: n equals a, t, c, or g
;   FEATURE:
;   NAME/KEY: misc_feature
;   LOCATION: (312837)..(312837)
;   OTHER INFORMATION: n equals a, t, c, or g
;   FEATURE:
;   NAME/KEY: misc_feature
;   LOCATION: (312993)..(312993)
;   OTHER INFORMATION: n equals a, t, c, or g
;   FEATURE:
;   NAME/KEY: misc_feature
;   LOCATION: (319226)..(319226)
;   OTHER INFORMATION: n equals a, t, c, or g
;   FEATURE:
;   NAME/KEY: misc_feature
;   LOCATION: (559167)..(559167)
;   OTHER INFORMATION: n equals a, t, c, or g
;   FEATURE:
;   NAME/KEY: misc_feature
;   LOCATION: (559241)..(559241)
;   OTHER INFORMATION: n equals a, t, c, or g
;   FEATURE:
;   NAME/KEY: misc_feature
;   LOCATION: (600992)..(600992)
;   OTHER INFORMATION: n equals a, t, c, or g
;   FEATURE:
;   NAME/KEY: misc_feature
;   LOCATION: (622708)..(622708)
;   OTHER INFORMATION: n equals a, t, c, or g
;   FEATURE:
;   NAME/KEY: misc_feature
;   LOCATION: (657081)..(657081)
;   OTHER INFORMATION: n equals a, t, c, or g
;   FEATURE:
;   NAME/KEY: misc_feature
;   LOCATION: (657203)..(657203)
;   OTHER INFORMATION: n equals a, t, c, or g
;   FEATURE:
;   NAME/KEY: misc_feature
;   LOCATION: (674435)..(674435)
;   OTHER INFORMATION: n equals a, t, c, or g
;   FEATURE:
;   NAME/KEY: misc_feature
;   LOCATION: (682442)..(682442)
;   OTHER INFORMATION: n equals a, t, c, or g
;   FEATURE:
;   NAME/KEY: misc_feature
;   LOCATION: (713652)..(713652)
;   OTHER INFORMATION: n equals a, t, c, or g
;   FEATURE:
;   NAME/KEY: misc_feature
;   LOCATION: (741684)..(741684)
;   OTHER INFORMATION: n equals a, t, c, or g
;   FEATURE:
;   NAME/KEY: misc_feature
;   LOCATION: (779455)..(779455)
;   OTHER INFORMATION: n equals a, t, c, or g
;   FEATURE:
;   NAME/KEY: misc_feature
;   LOCATION: (779676)..(779676)
;   OTHER INFORMATION: n equals a, t, c, or g
;   FEATURE:
;   NAME/KEY: misc_feature
;   LOCATION: (855539)..(855539)
;   OTHER INFORMATION: n equals a, t, c, or g
;   FEATURE:
;   NAME/KEY: misc_feature
;   LOCATION: (871619)..(871619)
;   OTHER INFORMATION: n equals a, t, c, or g
;   FEATURE:
;   NAME/KEY: misc_feature
;   LOCATION: (1084830)..(1084830)
;   OTHER INFORMATION: n equals a, t, c, or g
;   FEATURE:
;   NAME/KEY: misc_feature
;   LOCATION: (1096846)..(1096846)
;   OTHER INFORMATION: n equals a, t, c, or g
;   FEATURE:
;   NAME/KEY: misc_feature
;   LOCATION: (1119881)..(1119881)
;   OTHER INFORMATION: n equals a, t, c, or g
;   FEATURE:
;   NAME/KEY: misc_feature
;   LOCATION: (1130881)..(1130881)
;   OTHER INFORMATION: n equals a, t, c, or g
;   FEATURE:
;   NAME/KEY: misc_feature
;   LOCATION: (1310988)..(1310988)
;   OTHER INFORMATION: n equals a, t, c, or g
;   FEATURE:
;   NAME/KEY: misc_feature
;   LOCATION: (1313224)..(1313224)
;   OTHER INFORMATION: n equals a, t, c, or g
;   FEATURE:
;   NAME/KEY: misc_feature
;   LOCATION: (1349473)..(1349473)
;   OTHER INFORMATION: n equals a, t, c, or g

Alignment Scores:
Pred. No.:               1.19e-40      Length:         1664976
Score:                   656.00        Matches:        202
Percent Similarity:      46.75%        Conservative:   143
Best Local Similarity:   27.37%        Mismatches:     258
Query Match:             14.23%        Indels:         135
DB:                      3             Gaps:           26

US-10-768-511-8 (1-901) x US-09-692-570-1 (1-1664976)

Qy        177 TyrValTrpGly-ThrAsnValAsnIleProAsp---------ValLeuArgAlaIleAr 193
              :::|||||||||   ::::::  :::  |||          :::|||    :::||
Db      10758 TTCGTATGGGGATGTAGATATGGAATTGAGAGATGAAGACTTAATTTTAGAGGAAGTTAG 10699

Qy        193 gArgPheLeuHisAsnTyrArgSerSerAlaHis------------------------- 204
              |   :::|||     |||    :::  |||
Db      10698 GGATTATTTAACTGCCTATTTAAGAAATATTCATCAGGAAGATATTATCTTAGACAATGA 10639

Qy        205 ------------AspLeuAsnSerLysTyr--------------------IleGlnIl 213
                          |||||||||   |||                     ::::::
Db      10638 AAGAGTTGTAGTTGATTTAAATCAACTCTACAATTACGGATTAATGGAATTTGTAGAATT 10579

Qy        213 eIleGluGluThrValGluArgGluGluAspThrLeuAsnIleAspMetSerAspIleTy 233
               :::           ::::::          |||   :::            :::|||  ||
Db      10578 TTTAATAAATAATCCACAAAAAGGGATTGATTTTATAAAAGAGTGCTACAACGATGCTTA 10519

Qy        233 rAspHisAspProAspLeuTyrAlaLysIleValArgTyrProLeuAspIleIleProLe 253
              |                  |||              ||||||    ::::::|||
Db      10518 T-----------------TACACCTTAAGAAATGAATATCCAACAAACGTAATA----- 10482

Qy        253 uLeuAspThrGluCysGlnGluValAlaThrSerLeuLeuPro---ThrPheGluLysHi 272
                                    :::|||            ||||||     |||:::
Db      10481 ----------------------ATTGCTGTAAAAAATCTCCCAAAAATTTTTAAAACTAC 10444

Qy        272 sIleGluAlaArgProPheAsnLeuLysAlaSerValHisMetArgGluLeuAsnProSe 292
                 :::    :::   |||  ::::::                  ::::::
Db      10443 AAGAAAAGGGAAAATTTTCACAATAGAG------------------GATATTAGGAGCAA 10402
```

```
Qy       292 rAspIleAspLysLeuValSerValLysGlyMetValIleArgCysSerSerIleIlePr 312
Db     10401 AACTTTAGGAAAATTAGTTGAATTTGAAGGAATCATTGTTATGGCATCAAAAATAAGACC 10342

Qy       312 oGluIleLysGlyAlaPhePheLysCysLeuValCysGlyHisSerProProLeuValTh 332
Db     10341 AATGCTTAAAAAAGCTTATTATATATGTCCAAAATGTGGAAGAGAA-------------- 10296

Qy       332 rValValLysGlyArgValGluGluProThrArgCysGluLysProGluCysAlaAlaAr 352
Db     10295 -GTTGTTAGAGAAATAGATATCTTAAACACTGACTCTGAAAAAGCTGTTTGTGAATGTGG 10237

Qy       352 gAsnAlaMetSerLeuIleHisAsnArgCysThrPheAlaAsnLysGlnIleValArgLe 372
Db     10236 AGCTGAGTTAAACCTAATTGAGCATAAATCAATATATACTGACTTCCAAGAGATTAAAGT 10177

Qy       372 uGlnGluThrProAspAlaIle------ProGluGlyGluThrProHisThrValSerMe 390
Db     10176 TCAACAGCCATTGGATTTAATGGAAAATCCTGAA---GAGCCGCCAAAATACATAACTGT 10120

Qy       390 tCysLeuTyrAsnThrMetValAspAlaValLysProGly---------AspArgIleGl 407
Db     10119 GTTCTTAGAAAACAGT------------------CCTGGAATATATGCTGGAAGGGTGAA 10078

Qy       407 uValThrGlyValPheLysAlaMetAlaValArgValGlyProAsnGlnArgThrLeuAr 427
Db     10077 GATAACTGGCATC------------CCAATAAAAGTTAAGAAGAGTAAAAAGCTA----- 10035

Qy       427 gAlaLeuTyrLysThrTyrIleAspCysValHisValLysLysSerAspArgGlyArgLe 447
Db     10034 -CCAATCTATGACATACATGTTAAAGCCCTACACTGTGAGGTTTTGGAT----------- 9987

Qy       447 uGlnThrGluAspProMetGluMetAspLysGluAsnAspMetTyrAlaGlyTyrHisGl 467
Db      9987 ------------------------------------------------------------ 9987

Qy       467 uSerAspThrSerGluAlaAlaAsnGluAlaLysIleGlnLysLeuLysGluLeuSerLy 487
Db      9986 -GGGGAAGTAAAAATTAAATTAACAAATTCAGATATTGAAAATATTAAAAAAATAGCTAA 9928

Qy       487 sLeuProGlyIleTyrAspArgLeuSerArgSerLeuAlaProSerIleTrpGluLeuGl 507
Db      9927 AAGAAAAGATGTTGTTAATATACTTGCAGATAGATTAATTCCAGAGATTAAGGGGCATTC 9868

Qy       507 uAspIleLysLysGlyLeuLeuCysGlnLeuPheGlyGlyLysAlaLysLysIleProSe 527
Db      9867 TGCAATAAAGAAGGCTGTCTTACTACAACAGATAAAGGGAGTTAAAAAA----------- 9819

Qy       527 rGlyAlaSerPheArgGlyAspIleAsnValLeuLeuValGlyAspProGlyThrSerLy 547
Db      9818 ----CCTGGAAAGAGGGCTGATATTCATATATTATTAATCACAGACCCAGGAATTGGAAA 9763

Qy       547 sSerGlnLeuLeuGlnTyrValHisLysIleAlaProArgGlyIleTyrThrSerGlyAr 567
Db      9762 AACAGTCATACTAAGAAAGATTGCTGAAATC---CCAGGAAATTTATACGGTTCTGTTAC 9706

Qy       567 gGlySerSerAlaValGlyLeuThrAlaTyrValThrLysAsp---ProGluThrArgGl 586
Db      9705 TACCGCCACTGGTGTTGGGCTAACTGCCGCTGTTGTTAGGGACAAGACAGAGATTGGTGA 9646

Qy       586 uThr------ValLeuGluSerGlyAlaLeuValLeuSerAspArgGlyIleCysCysIl 604
Db      9645 AGATACATGGGTTATTAAGCCAGGTTTGTTAGTTAAAGCCCACAAAGGAACTGCATGTAT 9586

Qy       604 eAspGluPheAspLysMetSerAspAsnAlaArgSerMetLeuHisGluValMetGluGl 624
Db      9585 AGATGAGCTA---ACTGTTAATAAAGAACTGCAGAGCTATGTTTTGGAGGCTATGGAGAG 9529

Qy       624 nGlnThrValSerValAlaLysGlyGlyIleIleAlaSerLeuAsnAlaArgThrSerVa 644
Db      9528 TCAGACAATCCATATTAGTAAAGGAGGAATAAATGCAAAACTTCCAGCTGAATGTGCAAT 9469

Qy       644 lLeuAlaCysAlaAsnProSerGlySerArgTyrAsnAlaArgLeuSerValIleAspAs 664
Db      9468 TTTAGCCGCTTGCAATCCAAGATGGGGAAGATTCAACCCAGAGGTTTCTGTAGCTGAGCA 9409

Qy       664 nIleGlnLeuProProThrLeuLeuSerArgPheAspLeuIleTyrLeuMetLeuAspLy 684
Db      9408 GATAAATATTCCAGCCCCATTATTGAGTAGATTTGACCTAATATTTCCAATTAGAGATGT 9349

Qy       684 sProAspGluGlnAsnAspArgArgLeuAlaArgHisLeuValAlaLeuHis------Ty 702
Db      9348 TTCTGATAAGGATAAGGATAAAGATATCGCTGAATACATTGTAGATTTGCATAGAGCATA 9289

Qy       702 rGluAsnTyrGluValSerLysGlnAspAlaLeuAspLeu------------------- 715
Db      9288 TTTAGATGAGAAAATAAACAGGGAGATGGGTTTAGATTACCTTGAGGTTGATGGAGTTAA 9229

Qy       716 ----------GlnThrLeuThrAlaTyrIleThrTyrAlaArgGlnHisValHisProTh 732
Db      9228 AATAGACAAGGAATTTATAATAAAATATATCTACTACGCAAGACAGAAA---AAGCCGAT 9172

Qy       732 rLeuSerAspGluAlaAlaGluAspLeuIleAsnGlyTyrValGluMetArgGlnLysGl 752
Db      9171 AATTAGTGAAAAGGCTAAGGAATTATTTGTTAATTATTATGTAGAGATGAGAAAAAAACA 9112

Qy       752 yAsnPheProGlySerSerLysLysValIleThrAlaThrProArgGlnLeuGluSerMe 772
Db      9111 TCAGATA-----------------------------ACTGCAAGGCAGTTAGAGGCTGC 9082

Qy       772 tIleArgIleSerGluAlaLeuAlaArgMetArgPheSerGluValValGluLysValAs 792
Db      9081 TATAAGGATTGCTGAAGCTCATGCAAAGGCAAAGTTAAAGGATGTTGTTGATGAGGAAGA 9022

Qy       792 pAlaAlaGluAlaValArgLeuLeuAspValAlaLeuGlnGlnSerAlaThrAspHisAl 812
Db      9021 TGCAAAGGAGGCAATAAATATAATAACTGAATGCTTAAAAGAAATTGCCTATGACCCAGA 8962

Qy       812 aThrGlyThrIleAspMetAspLeuIleThrThrGlyValSerAlaSerGluArgIleAr 832
Db      8961 GACTGGAATATTTGATGTTGATAAAATT---TTAGGGGTGTCTAAGAAAGAGAGAGACAA 8905

Qy       832 gArgAlaAsnLeuLeuAlaAlaLeuArgGluLeuIle---------------------- 844
Db      8904 ATTAACAACCGTCTATGAGATAATTAAGGAGTTATCTGAAAAATCAGAACTTGTTGAACA 8845

Qy       845 -AlaAspLysIleSerProGlySerSerSerGlyLeuLysThrSerGlnLeu 861
Db      8844 TGAAGATATAGCTGAAGAAGCAAAAAAGAAAGGTATTAAAGAGGATGAGTTG 8793
```

Search completed: December 6, 2005, 15:04:27
Job time : 4036 secs

```
OM protein - nucleic search, using frame_plus_p2n model

Run on:          December  6, 2005, 09:59:22 ; Search time 1136 Seconds
                                           (without alignments)
                                           6558.723 Million cell updates/sec

Title:           US-10-768-511-8
Perfect score:   4609
Sequence:        1 MENNDALDIGAVSSPYPSQS..........LGSLQGEGFLTVHGDIVKRV 901

Scoring table:   BLOSUM62
                 Xgapop 10.0 , Xgapext  0.5
                 Ygapop 10.0 , Ygapext  0.5
                 Fgapop  6.0 , Fgapext  7.0
                 Delop   6.0 , Delext   7.0

Searched:        9793542 seqs, 4134689005 residues

Total number of hits satisfying chosen parameters:      19587084

Minimum DB seq length: 0
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%
                 Maximum Match 100%
                 Listing first 45 summaries

Command line parameters:
-MODEL=frame+_p2n.model -DEV=xlh
-Q=/cgn2_1/USPTO_spool/US10768511/runat_02122005_102335_18467/app_query.fasta_1.1095
-DB=Published_Applications_NA_Main -QFMT=fastap -SUFFIX=rnpbm -MINMATCH=0.1
-LOOPCL=0 -LOOPEXT=0 -UNITS=bits -START=1 -END=-1 -MATRIX=blosum62
-TRANS=human40.cdi -LIST=45 -DOCALIGN=200 -THR_SCORE=pct -THR_MAX=100
-THR_MIN=0 -ALIGN=15 -MODE=LOCAL -OUTFMT=pto -NORM=ext -HEAPSIZE=500 -MINLEN=0
-MAXLEN=2000000000 -USER=US10768511_@CGN_1_1_797_@runat_02122005_102335_18467
-NCPU=6 -ICPU=3 -NO_MMAP -LARGEQUERY -NEG_SCORES=0 -WAIT -DSPBLOCK=100
-LONGLOG -DEV_TIMEOUT=120 -WARN_TIMEOUT=30 -THREADS=1 -XGAPOP=10 -XGAPEXT=0.5
-FGAPOP=6 -FGAPEXT=7 -YGAPOP=10 -YGAPEXT=0.5 -DELOP=6 -DELEXT=7

Database :       Published_Applications_NA_Main:*
                 1:  /cgn2_6/ptodata/1/pubpna/US07_PUBCOMB.seq:*
                 2:  /cgn2_6/ptodata/1/pubpna/US08_PUBCOMB.seq:*
                 3:  /cgn2_6/ptodata/1/pubpna/US09A_PUBCOMB.seq:*
                 4:  /cgn2_6/ptodata/1/pubpna/US09B_PUBCOMB.seq:*
                 5:  /cgn2_6/ptodata/1/pubpna/US10A_PUBCOMB.seq:*
                 6:  /cgn2_6/ptodata/1/pubpna/US10B_PUBCOMB.seq:*
                 7:  /cgn2_6/ptodata/1/pubpna/US10C_PUBCOMB.seq:*
                 8:  /cgn2_6/ptodata/1/pubpna/US10D_PUBCOMB.seq:*
                 9:  /cgn2_6/ptodata/1/pubpna/US10E_PUBCOMB.seq:*
                 10:  /cgn2_6/ptodata/1/pubpna/US11_PUBCOMB.seq:*

      Pred. No. is the number of results predicted by chance to have a
      score greater than or equal to the score of the result being printed,
      and is derived by analysis of the total score distribution.
```

SUMMARIES

Result No.	Score	% Query Match	Length	DB	ID	Description
1	4609	100.0	4348	3	US-09-828-062-5	Sequence 5, Appli
2	4609	100.0	4348	7	US-10-768-511-5	Sequence 5, Appli
3	2550	55.3	2841	7	US-10-425-114-28735	Sequence 28735, A
4	2481.5	53.8	3054	7	US-10-437-963-44639	Sequence 44639, A
5	1963.5	42.6	2186	8	US-10-425-115-98898	Sequence 98898, A
6	1883	40.9	3072	5	US-10-128-714-7558	Sequence 7558, Ap
7	1871.5	40.6	2625	5	US-10-128-714-2558	Sequence 2558, Ap
8	1854	40.2	2939	8	US-10-425-115-98897	Sequence 98897, A
9	1840	39.9	1723	7	US-10-260-238-580	Sequence 580, App
10	1835.5	39.8	2970	7	US-10-320-797-2332	Sequence 2332, Ap
11	1833	39.8	2802	6	US-10-369-493-46198	Sequence 46198, A
12	1819	39.5	2881	6	US-10-369-493-27488	Sequence 27488, A
13	1817.5	39.4	3273	3	US-09-954-456-568	Sequence 568, App
14	1817.5	39.4	3273	3	US-09-954-456-569	Sequence 569, App
15	1817.5	39.4	3273	3	US-09-954-456-976	Sequence 976, App
16	1817.5	39.4	3273	6	US-10-172-118-1920	Sequence 1920, Ap
17	1817.5	39.4	3273	7	US-10-342-887-1920	Sequence 1920, Ap
18	1817.5	39.4	3273	9	US-10-843-641A-3595	Sequence 3595, Ap
19	1817.5	39.4	3273	9	US-10-843-641A-3596	Sequence 3596, Ap
20	1817.5	39.4	3273	9	US-10-843-641A-4003	Sequence 4003, Ap
21	1817.5	39.4	3273	9	US-10-756-149-1918	Sequence 1918, Ap
22	1816.5	39.4	3395	5	US-10-084-817-287	Sequence 287, App
23	1791	38.9	3394	6	US-10-240-965-223	Sequence 223, App
24	1789.5	38.8	2842	3	US-09-925-300-113	Sequence 113, App
25	1764	38.3	3362	5	US-10-128-714-6558	Sequence 6558, Ap
26	1764	38.3	4944	5	US-10-128-714-558	Sequence 558, App
27	1764	38.3	5362	5	US-10-128-714-5558	Sequence 5558, Ap
28	1752.5	38.0	2944	5	US-10-128-714-1558	Sequence 1558, Ap
29	1752	38.0	2733	6	US-10-032-585-6480	Sequence 6480, Ap
30	1690.5	36.7	2692	10	US-11-097-143-1688	Sequence 1688, Ap
31	1677	36.4	5419	7	US-10-320-797-332	Sequence 332, App
32	1675	36.3	3419	7	US-10-320-797-1332	Sequence 1332, Ap
33	1670.5	36.2	1571	7	US-10-425-114-19722	Sequence 19722, A
34	1584.5	34.4	5065	10	US-11-097-143-1687	Sequence 1687, Ap
35	1139	24.7	723	3	US-09-828-062-2	Sequence 2, Appli
36	1139	24.7	723	7	US-10-768-511-2	Sequence 2, Appli
37	1021.5	22.2	2001	6	US-10-369-493-45225	Sequence 45225, A
38	1016.5	22.1	2280	6	US-10-369-493-46251	Sequence 46251, A
39	1016	22.0	2821	8	US-10-733-878-462	Sequence 462, App
40	1003	21.8	2445	5	US-10-128-714-7100	Sequence 7100, Ap
41	998	21.7	2427	3	US-09-880-107-1686	Sequence 1686, Ap
42	998	21.7	2427	3	US-09-873-367C-977	Sequence 977, App
43	998	21.7	2427	6	US-10-172-118-368	Sequence 368, App
44	998	21.7	2427	7	US-10-342-887-368	Sequence 368, App
45	998	21.7	2427	9	US-10-843-641A-977	Sequence 977, App

ALIGNMENTS

```
RESULT 1
US-09-828-062-5
; Sequence 5, Application US/09828062
; Publication No. US20030097675A1
; GENERAL INFORMATION:
;  APPLICANT: COSTA E SILVA, OSWALDO DA
;  APPLICANT:  BOHNERT, HANS J.
;  APPLICANT:  VAN THIELEN, NOCHA
;  APPLICANT:  CHEN, ROUYING
;  APPLICANT:  SARRIA-MILLAN, RODRIGO
;  TITLE OF INVENTION: CELL CYCLE STRESS-RELATED PROTEINS AND METHODS OF USE
;  TITLE OF INVENTION:  IN PLANTS
;  FILE REFERENCE: 16313-0031
;  CURRENT APPLICATION NUMBER: US/09/828,062
;  CURRENT FILING DATE:  2001-08-20
;  PRIOR APPLICATION NUMBER: 60/196,001
;  PRIOR FILING DATE: 2000-04-07
;  NUMBER OF SEQ ID NOS: 34
;  SOFTWARE: PatentIn Ver. 2.1
; SEQ ID NO 5
;   LENGTH: 4348
;   TYPE: DNA
;   ORGANISM: Physcomitrella patens
US-09-828-062-5

Alignment Scores:
Pred. No.:              0                 Length:          4348
Score:                  4609.00           Matches:         901
Percent Similarity:     100.00%           Conservative:    0
Best Local Similarity:  100.00%           Mismatches:      0
Query Match:            100.00%           Indels:          0
DB:                     3                 Gaps:            0
```

```
US-10-768-511-8 (1-901) x US-09-828-062-5 (1-4348)

Qy          1 MetGluAsnAsnAspAlaLeuAspIleGlyAlaValSerSerProTyrProSerGlnSer 20
                ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1554 ATGGAAAATAATGATGCACTTGACATTGGAGCCGTGTCGTCCCCATATCCTTCGCAATCT 1613

Qy         21 GluGlyValSerThrProLeuProGlnValThrSerProSerPheAspAsnAlaAlaSer 40
                ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1614 GAAGGAGTGTCTACGCCATTGCCGCAAGTAACATCACCGAGCTTCGACAATGCAGCCTCA 1673

Qy         41 ProValAlaGlyArgArgAlaValArgGlnThrProThrSerAlaValArgArgArgGly 60
                ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1674 CCCGTGGCCGGGCGGAGGGCCGTACGGCAGACCCCTACATCTGCAGTTCGAAGGAGAGGG 1733

Qy         61 ArgGluThrAspSerAlaArgArgArgSerArgSerArgSerLeuGlyAsnSerVal 80
                ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1734 AGAGAAACGGATTCCGCTCGTCGTAGGAGGAGTCGATCTCGCAGTTTAGGCAATTCTGTT 1793

Qy         81 TyrSerSerProTyrAspAlaGlyThrProGlyThrProGlyThrProValAlaThrPro 100
                ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1794 TATAGTTCCCCTTACGATGCGGGGACTCCTGGAACTCCTGGAACTCCAGTGGCTACTCCG 1853

Qy        101 ValTyrAlaThrProValGlyThrProMetGlyThrProSerPheHisArgGlyThrPro 120
                ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1854 GTTTACGCTACCCCAGTCGGTACACCTATGGGTACCCCATCGTTCCATCGTGGCACGCCA 1913

Qy        121 GlnTyrLysGlnArgSerGluLeuGlySerGlnGlyLysProLeuHisArgArgArgArg 140
                ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1914 CAGTACAAACAGCGCAGTGAGCTTGGTTCCCAGGGGAAGCCTCTACATCGGAGACGTCGA 1973

Qy        141 SerGlnSerArgGluProGlyHisArgSerProSerArgGluProSerAlaAspGlyArg 160
                ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1974 TCTCAATCCAGAGAACCCGGGCATCGATCTCCTTCAAGGGAACCTAGTGCTGATGGGCGT 2033

Qy        161 ProSerGluSerAlaGluProAspAspThrLeuGlyGlyGluTyrAlaTyrValTrpGly 180
                ||||||||||||||||||||||||||||||||||||||||||| ||||||||||||||||
Db       2034 CCCTCTGAATCTGCTGAGCCAGATGACACTTTGGGTGGAGAATATGCTTATGTTTGGGGG 2093

Qy        181 ThrAsnValAsnIleProAspValLeuArgAlaIleArgArgPheLeuHisAsnTyrArg 200
                ||||||||||||||||||||||||||||||||||||||||||| ||||||||||||||||
Db       2094 ACGAATGTTAACATTCCAGATGTGCTTAGGGCGATTCGTCGATTCTCCACAATTATCGT 2153

Qy        201 SerSerAlaHisAspLeuAsnSerLysTyrIleGlnIleIleGluGluThrValGluArg 220
                ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       2154 TCGAGTGCTCATGATCTTAATTCCAAGTACATCCAGATCATAGAGGAGACTGTGGAGCGT 2213

Qy        221 GluGluAspThrLeuAsnIleAspMetSerAspIleTyrAspHisAspProAspLeuTyr 240
                ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       2214 GAGGAGGATACTCTAAATATCGACATGTCAGACATTTATGACCATGATCCTGATCTATAC 2273

Qy        241 AlaLysIleValArgTyrProLeuAspIleIleProLeuLeuAspThrGluCysGlnGlu 260
                ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       2274 GCAAAAATTGTTCGATACCCACTCGACATCATCCCCCTGTTGGACACTGAGTGTCAGGAA 2333

Qy        261 ValAlaThrSerLeuLeuProThrPheGluLysHisIleGluAlaArgProPheAsnLeu 280
                ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       2334 GTTGCTACCTCTTTACTACCAACGTTTGAGAAGCATATTGAGGCCAGACCTTTCAATCTC 2393

Qy        281 LysAlaSerValHisMetArgGluLeuAsnProSerAspIleAspLysLeuValSerVal 300
                ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       2394 AAAGCATCGGTGCACATGCGTGAACTCAACCCTTCAGATATAGACAAATTGGTTTCTGTT 2453

Qy        301 LysGlyMetValIleArgCysSerSerIleIleProGluIleLysGlyAlaPhePheLys 320
                ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       2454 AAAGGAATGGTTATCCGGTGCAGTTCTATCATACCTGAAATTAAGGGGGCCTTCTTCAAA 2513

Qy        321 CysLeuValCysGlyHisSerProProLeuValThrValValLysGlyArgValGluGlu 340
                ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       2514 TGTTTAGTGTGTGGTCACTCGCCTCCGCTAGTTACAGTTGTTAAAGGGCGGGTTGAGGAG 2573

Qy        341 ProThrArgCysGluLysProGluCysAlaAlaArgAsnAlaMetSerLeuIleHisAsn 360
                ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       2574 CCAACAAGGTGTGAAAAGCCAGAATGTGCAGCACGGAATGCTATGTCTCTTATTCACAAT 2633

Qy        361 ArgCysThrPheAlaAsnLysGlnIleValArgLeuGlnGluThrProAspAlaIlePro 380
                ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       2634 CGATGCACTTTTGCAAATAAGCAGATAGTGCGTCTTCAAGAAACTCCAGATGCCATTCCT 2693

Qy        381 GluGlyGluThrProHisThrValSerMetCysLeuTyrAsnThrMetValAspAlaVal 400
                ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       2694 GAAGGAGAGACTCCACACACAGTCAGCATGTGTTTATACAACACTATGGTTGATGCTGTG 2753

Qy        401 LysProGlyAspArgIleGluValThrGlyValPheLysAlaMetAlaValArgValGly 420
                ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       2754 AAGCCTGGAGATCGTATTGAGGTAACAGGAGTTTTCAAGGCCATGGCAGTTCGAGTTGGT 2813

Qy        421 ProAsnGlnArgThrLeuArgAlaLeuTyrLysThrTyrIleAspCysValHisValLys 440
                ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       2814 CCGAATCAACGAACATTACGAGCATTGTATAAGACCTACATCGATTGCGTGCACGTCAAG 2873

Qy        441 LysSerAspArgGlyArgLeuGlnThrGluAspProMetGluMetAspLysGluAsnAsp 460
                ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       2874 AAGTCTGACAGGGGTCGACTGCAAACTGAAGATCCTATGGAGATGGATAAGGAGAATGAT 2933

Qy        461 MetTyrAlaGlyTyrHisGluSerAspThrSerGluAlaAlaAsnGluAlaLysIleGln 480
                ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       2934 ATGTATGCTGGGTATCATGAAAGTGATACTTCAGAAGCTGCTAATGAAGCAAAGATTCAA 2993

Qy        481 LysLeuLysGluLeuSerLysLeuProGlyIleTyrAspArgLeuSerArgSerLeuAla 500
                ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       2994 AAACTTAAAGAGCTGTCCAAGCTCCCGGGCATTTATGATAGACTTTCAAGGTCGCTGGCT 3053

Qy        501 ProSerIleTrpGluLeuGluAspIleLysLysGlyLeuLeuCysGlnLeuPheGlyGly 520
                ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       3054 CCAAGCATTTGGGAGCTTGAAGATATTAAAAAGGGTCTTCTTTGCCAGCTCTTTGGTGGG 3113

Qy        521 LysAlaLysLysIleProSerGlyAlaSerPheArgGlyAspIleAsnValLeuLeuVal 540
                ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       3114 AAGGCTAAGAAAATTCCATCTGGAGCATCTTTCCGAGGTGACATCAATGTTTTACTTGTT 3173

Qy        541 GlyAspProGlyThrSerLysSerGlnLeuLeuGlnTyrValHisLysIleAlaProArg 560
                ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       3174 GGGGACCCTGGTACCAGTAAATCTCAGCTGCTTCAGTATGTGCACAAGATAGCTCCTCGT 3233

Qy        561 GlyIleTyrThrSerGlyArgGlySerSerAlaValGlyLeuThrAlaTyrValThrLys 580
                ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       3234 GGAATCTACACTAGTGGGCGAGGAAGTTCGGCGGTTGGGCTGACAGCGTATGTAACGAAG 3293

Qy        581 AspProGluThrArgGluThrValLeuGluSerGlyAlaLeuValLeuSerAspArgGly 600
                ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       3294 GATCCAGAAACTCGAGAGACGGTATTGGAGAGCGGAGCTTTGGTTCTTAGTGATCGTGGG 3353

Qy        601 IleCysCysIleAspGluPheAspLysMetSerAspAsnAlaArgSerMetLeuHisGlu 620
                ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       3354 ATATGCTGTATCGATGAGTTCGACAAAATGTCTGATAATGCCCGAAGCATGCTTCATGAG 3413

Qy        621 ValMetGluGlnGlnThrValSerValAlaLysGlyGlyIleIleAlaSerLeuAsnAla 640
                ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       3414 GTAATGGAGCAACAAACGGTATCTGTAGCCAAAGGGGGTATCATTGCCTCGCTGAACGCT 3473

Qy        641 ArgThrSerValLeuAlaCysAlaAsnProSerGlySerArgTyrAsnAlaArgLeuSer 660
                ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       3474 CGGACGTCTGTCCTTGCATGTGCAAATCCTAGTGGGTCCCGATACAATGCGCGCCTTTCT 3533

Qy        661 ValIleAspAsnIleGlnLeuProProThrLeuLeuSerArgPheAspLeuIleTyrLeu 680
                ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       3534 GTGATTGATAACATCCAGCTTCCTCCAACTCTACTTTCTAGATTTGATTTAATTTACTTA 3593

Qy        681 MetLeuAspLysProAspGluGlnAsnAspArgArgLeuAlaArgHisLeuValAlaLeu 700
                ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       3594 ATGCTCGACAAACCAGACGAGCAAAACGATCGTCGTCTCGCCAGGCATCTCGTGGCTTTA 3653

Qy        701 HisTyrGluAsnTyrGluValSerLysGlnAspAlaLeuAspLeuGlnThrLeuThrAla 720
                ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       3654 CACTATGAAAACTATGAAGTTTCAAAGCAGGACGCCTTAGATCTACAAACACTTACCGCG 3713
```

```
Qy        721 TyrIleThrTyrAlaArgGlnHisValHisProThrLeuSerAspGluAlaAlaGluAsp 740
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       3714 TATATCACCTATGCTCGTCAGCATGTACATCCTACATTAAGTGATGAAGCTGCTGAAGAT 3773

Qy        741 LeuIleAsnGlyTyrValGluMetArgGlnLysGlyAsnPheProGlySerSerLysLys 760
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       3774 TTGATTAATGGCTATGTTGAGATGCGCCAAAAGGGCAACTTTCCTGGAAGCAGTAAAAAG 3833

Qy        761 ValIleThrAlaThrProArgGlnLeuGluSerMetIleArgIleSerGluAlaLeuAla 780
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       3834 GTGATAACAGCCACACCTCGGCAACTCGAAAGTATGATTCGTATCAGTGAAGCCCTAGCT 3893

Qy        781 ArgMetArgPheSerGluValValGluLysValAspAlaAlaGluAlaValArgLeuLeu 800
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       3894 CGAATGAGATTTTCTGAAGTGGTAGAGAAAGTTGATGCAGCAGAAGCTGTGCGCCTTTTA 3953

Qy        801 AspValAlaLeuGlnGlnSerAlaThrAspHisAlaThrGlyThrIleAspMetAspLeu 820
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       3954 GACGTCGCTTTGCAGCAATCTGCTACTGATCATGCAACAGGTACGATAGACATGGATCTT 4013

Qy        821 IleThrThrGlyValSerAlaSerGluArgIleArgArgAlaAsnLeuLeuAlaAlaLeu 840
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       4014 ATCACGACTGGAGTGTCGGCCAGCGAGCGTATTCGTCGGGCCAACTTGCTAGCTGCTCTG 4073

Qy        841 ArgGluLeuIleAlaAspLysIleSerProGlySerSerSerGlyLeuLysThrSerGln 860
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       4074 CGAGAGCTTATAGCAGATAAAATTTCACCTGGCAGCTCCTCTGGCTTGAAGACCAGTCAG 4133

Qy        861 LeuLeuGluAspIleArgSerGlnSerSerValAspValSerLeuGlnAspIleLysAsn 880
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       4134 CTTCTTGAGGATATCCGGAGCCAAAGCAGTGTGGACGTTAGTTTGCAGGATATTAAAAAT 4193

Qy        881 AlaLeuGlySerLeuGlnGlyGluGlyPheLeuThrValHisGlyAspIleValLysArg 900
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       4194 GCTCTGGGTAGCCTCCAAGGAGAAGGCTTTCTTACTGTCCATGGTGACATAGTCAAGAGA 4253

Qy        901 Val 901
              |||
Db       4254 GTT 4256
```

```
RESULT 2
US-10-768-511-5
; Sequence 5, Application US/10768511
; Publication No. US20040128721A1
; GENERAL INFORMATION:
;  APPLICANT: COSTA E SILVA, OSWALDO DA
;  APPLICANT:  BOHNERT, HANS J.
;  APPLICANT:  VAN THIELEN, NOCHA
;  APPLICANT:  CHEN, ROUYING
;  APPLICANT:  SARRIA-MILLAN, RODRIGO
;  TITLE OF INVENTION: CELL CYCLE STRESS-RELATED PROTEINS AND METHODS OF USE
;  TITLE OF INVENTION:  IN PLANTS
;  FILE REFERENCE: 16313-0031
;  CURRENT APPLICATION NUMBER: US/10/768,511
;  CURRENT FILING DATE:  2004-01-30
;  PRIOR APPLICATION NUMBER: 60/196,001
;  PRIOR FILING DATE: 2000-04-07
;  NUMBER OF SEQ ID NOS: 34
;  SOFTWARE: PatentIn Ver. 2.1
; SEQ ID NO 5
;   LENGTH: 4348
;   TYPE: DNA
;   ORGANISM: Physcomitrella patens
US-10-768-511-5

Alignment Scores:
Pred. No.:              0              Length:         4348
Score:                  4609.00        Matches:        901
Percent Similarity:     100.00%        Conservative:   0
Best Local Similarity:  100.00%        Mismatches:     0
Query Match:            100.00%        Indels:         0
DB:                     7              Gaps:           0
```

```
US-10-768-511-8 (1-901) x US-10-768-511-5 (1-4348)

Qy          1 MetGluAsnAsnAspAlaLeuAspIleGlyAlaValSerSerProTyrProSerGlnSer 20
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1554 ATGGAAAATAATGATGCACTTGACATTGGAGCCGTGTCGTCCCCATATCCTTCGCAATCT 1613

Qy         21 GluGlyValSerThrProLeuProGlnValThrSerProSerPheAspAsnAlaAlaSer 40
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1614 GAAGGAGTGTCTACGCCATTGCCGCAAGTAACATCACCGAGCTTCGACAATGCAGCCTCA 1673

Qy         41 ProValAlaGlyArgArgAlaValArgGlnThrProThrSerAlaValArgArgArgGly 60
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1674 CCCGTGGCCGGGCGGAGGGCCGTACGGCAGACCCCTACATCTGCAGTTCGAAGGAGAGGG 1733

Qy         61 ArgGluThrAspSerAlaArgArgArgArgSerArgSerArgSerLeuGlyAsnSerVal 80
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1734 AGAGAAACGGATTCCGCTCGTCGTAGGAGGAGTCGATCTCGCAGTTTAGGCAATTCTGTT 1793

Qy         81 TyrSerSerProTyrAspAlaGlyThrProGlyThrProGlyThrProValAlaThrPro 100
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1794 TATAGTTCCCCTTACGATGCGGGGACTCCTGGAACTCCTGGAACTCCAGTGGCTACTCCG 1853

Qy        101 ValTyrAlaThrProValGlyThrProMetGlyThrProSerPheHisArgGlyThrPro 120
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1854 GTTTACGCTACCCCAGTCGGTACACCTATGGGTACCCCATCGTTCCATCGTGGCACGCCA 1913

Qy        121 GlnTyrLysGlnArgSerGluLeuGlySerGlnGlyLysProLeuHisArgArgArgArg 140
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1914 CAGTACAAACAGCGCAGTGAGCTTGGTTCCCAGGGGAAGCCTCTACATCGGAGACGTCGA 1973

Qy        141 SerGlnSerArgGluProGlyHisArgSerProSerArgGluProSerAlaAspGlyArg 160
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1974 TCTCAATCCAGAGAACCCGGGCATCGATCTCCTTCAAGGGAACCTAGTGCTGATGGGCGT 2033

Qy        161 ProSerGluSerAlaGluProAspAspThrLeuGlyGlyGluTyrAlaTyrValTrpGly 180
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       2034 CCCTCTGAATCTGCTGAGCCAGATGACACTTTGGGTGGAGAATATGCTTATGTTTGGGGG 2093

Qy        181 ThrAsnValAsnIleProAspValLeuArgAlaIleArgArgPheLeuHisAsnTyrArg 200
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       2094 ACGAATGTTAACATTCCAGATGTGCTTAGGGCGATTCGTCGATTTCTCCACAATTATCGT 2153

Qy        201 SerSerAlaHisAspLeuAsnSerLysTyrIleGlnIleIleGluGluThrValGluArg 220
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       2154 TCGAGTGCTCATGATCTTAATTCCAAGTACATCCAGATCATAGAGGAGACTGTGGAGCGT 2213

Qy        221 GluGluAspThrLeuAsnIleAspMetSerAspIleTyrAspHisAspProAspLeuTyr 240
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       2214 GAGGAGGATACTCTAAATATCGACATGTCAGACATTTATGACCATGATCCTGATCTATAC 2273

Qy        241 AlaLysIleValArgTyrProLeuAspIleIleProLeuLeuAspThrGluCysGlnGlu 260
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       2274 GCAAAAATTGTTCGATACCCACTCGACATCATCCCCCTGTTGGACACTGAGTGTCAGGAA 2333

Qy        261 ValAlaThrSerLeuLeuProThrPheGluLysHisIleGluAlaArgProPheAsnLeu 280
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       2334 GTTGCTACCTCTTTACTACCAACGTTTGAGAAGCATATTGAGGCCAGACCTTTCAATCTC 2393

Qy        281 LysAlaSerValHisMetArgGluLeuAsnProSerAspIleAspLysLeuValSerVal 300
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       2394 AAAGCATCGGTGCACATGCGTGAACTCAACCCTTCAGATATAGACAAATTGGTTTCTGTT 2453

Qy        301 LysGlyMetValIleArgCysSerSerIleIleProGluIleLysGlyAlaPhePheLys 320
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       2454 AAAGGAATGGTTATCCGGTGCAGTTCTATCATACCTGAAATTAAGGGGGCCTTCTTCAAA 2513

Qy        321 CysLeuValCysGlyHisSerProProLeuValThrValValLysGlyArgValGluGlu 340
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       2514 TGTTTAGTGTGTGGTCACTCGCCTCCGCTAGTTACAGTTGTTAAAGGGCGGGTTGAGGAG 2573

Qy        341 ProThrArgCysGluLysProGluCysAlaAlaArgAsnAlaMetSerLeuIleHisAsn 360
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
```

```
Db   2574 CCAACAAGGTGTGAAAAGCCAGAATGTGCAGCACGGAATGCTATGTCTCTTATTCACAAT 2633
Qy    361 ArgCysThrPheAlaAsnLysGlnIleValArgLeuGlnGluThrProAspAlaIlePro 380
Db   2634 CGATGCACTTTTGCAAATAAGCAGATAGTGCGTCTTCAAGAAACTCCAGATGCCATTCCT 2693
Qy    381 GluGlyGluThrProHisThrValSerMetCysLeuTyrAsnThrMetValAspAlaVal 400
Db   2694 GAAGGAGAGACTCCACACACAGTCAGCATGTGTTTATACAACACTATGGTTGATGCTGTG 2753
Qy    401 LysProGlyAspArgIleGluValThrGlyValPheLysAlaMetAlaValArgValGly 420
Db   2754 AAGCCTGGAGATCGTATTGAGGTAACAGGAGTTTTCAAGGCCATGGCAGTTCGAGTTGGT 2813
Qy    421 ProAsnGlnArgThrLeuArgAlaLeuTyrLysThrTyrIleAspCysValHisValLys 440
Db   2814 CCGAATCAACGAACATTACGAGCATTGTATAAGACCTACATCGATTGCGTGCACGTCAAG 2873
Qy    441 LysSerAspArgGlyArgLeuGlnThrGluAspProMetGluMetAspLysGluAsnAsp 460
Db   2874 AAGTCTGACAGGGGTCGACTGCAAACTGAAGATCCTATGGAGATGGATAAGGAGAATGAT 2933
Qy    461 MetTyrAlaGlyTyrHisGluSerAspThrSerGluAlaAlaAsnGluAlaLysIleGln 480
Db   2934 ATGTATGCTGGGTATCATGAAAGTGATACTTCAGAAGCTGCTAATGAAGCAAAGATTCAA 2993
Qy    481 LysLeuLysGluLeuSerLysLeuProGlyIleTyrAspArgLeuSerArgSerLeuAla 500
Db   2994 AAACTTAAAGAGCTGTCCAAGCTCCCGGGCATTTATGATAGACTTTCAAGGTCGCTGGCT 3053
Qy    501 ProSerIleTrpGluLeuGluAspIleLysLysGlyLeuLeuCysGlnLeuPheGlyGly 520
Db   3054 CCAAGCATTTGGGAGCTTGAAGATATTAAAAAGGGTCTTCTTTGCCAGCTCTTTGGTGGG 3113
Qy    521 LysAlaLysLysIleProSerGlyAlaSerPheArgGlyAspIleAsnValLeuLeuVal 540
Db   3114 AAGGCTAAGAAAATTCCATCTGGAGCATCTTTCCGAGGTGACATCAATGTTTTACTTGTT 3173
Qy    541 GlyAspProGlyThrSerLysSerGlnLeuLeuGlnTyrValHisLysIleAlaProArg 560
Db   3174 GGGGACCCTGGTACCAGTAAATCTCAGCTGCTTCAGTATGTGCACAAGATAGCTCCTCGT 3233
Qy    561 GlyIleTyrThrSerGlyArgGlySerSerAlaValGlyLeuThrAlaTyrValThrLys 580
Db   3234 GGAATCTACACTAGTGGGCGAGGAAGTTCGGCGGTTGGGCTGACAGCGTATGTAACGAAG 3293
Qy    581 AspProGluThrArgGluThrValLeuGluSerGlyAlaLeuValLeuSerAspArgGly 600
Db   3294 GATCCAGAAACTCGAGAGACGGTATTGGAGAGCGGAGCTTTGGTTCTTAGTGATCGTGGG 3353
Qy    601 IleCysCysIleAspGluPheAspLysMetSerAspAsnAlaArgSerMetLeuHisGlu 620
Db   3354 ATATGCTGTATCGATGAGTTCGACAAAATGTCTGATAATGCCCGAAGCATGCTTCATGAG 3413
Qy    621 ValMetGluGlnGlnThrValSerValAlaLysGlyGlyIleIleAlaSerLeuAsnAla 640
Db   3414 GTAATGGAGCAACAAACGGTATCTGTAGCCAAGGGGGTATCATTGCCTCGCTGAACGCT 3473
Qy    641 ArgThrSerValLeuAlaCysAlaAsnProSerGlySerArgTyrAsnAlaArgLeuSer 660
Db   3474 CGGACGTCTGTCCTTGCATGTGCAAATCCTAGTGGGTCCCGATACAATGCGCGCCTTTCT 3533
Qy    661 ValIleAspAsnIleGlnLeuProProThrLeuLeuSerArgPheAspLeuIleTyrLeu 680
Db   3534 GTGATTGATAACATCCAGCTTCCTCCAACTCTACTTTCTAGATTTGATTTAATTTACTTA 3593
Qy    681 MetLeuAspLysProAspGluGlnAsnAspArgArgLeuAlaArgHisLeuValAlaLeu 700
Db   3594 ATGCTCGACAAACCAGACGAGCAAAACGATCGTCGTCTCGCCAGGCATCTCGTGGCTTTA 3653
Qy    701 HisTyrGluAsnTyrGluValSerLysGlnAspAlaLeuAspLeuGlnThrLeuThrAla 720
Db   3654 CACTATGAAAACTATGAAGTTTCAAAGCAGGACGCCTTAGATCTACAAACACTTACCGCG 3713
Qy    721 TyrIleThrTyrAlaArgGlnHisValHisProThrLeuSerAspGluAlaAlaGluAsp 740
Db   3714 TATATCACCTATGCTCGTCAGCATGTACATCCTACATTAAGTGATGAAGCTGCTGAAGAT 3773
Qy    741 LeuIleAsnGlyTyrValGluMetArgGlnLysGlyAsnPheProGlySerSerLysLys 760
Db   3774 TTGATTAATGGCTATGTTGAGATGCGCCAAAAGGGCAACTTTCCTGGAAGCAGTAAAAAG 3833
Qy    761 ValIleThrAlaThrProArgGlnLeuGluSerMetIleArgIleSerGluAlaLeuAla 780
Db   3834 GTGATAACAGCCACACCTCGGCAACTCGAAAGTATGATTCGTATCAGTGAAGCCCTAGCT 3893
Qy    781 ArgMetArgPheSerGluValValGluLysValAspAlaAlaGluAlaValArgLeuLeu 800
Db   3894 CGAATGAGATTTTCTGAAGTGGTAGAGAAAGTTGATGCAGCAGAAGCTGTGCGCCTTTTA 3953
Qy    801 AspValAlaLeuGlnGlnSerAlaThrAspHisAlaThrGlyThrIleAspMetAspLeu 820
Db   3954 GACGTCGCTTTGCAGCAATCTGCTACTGATCATGCAACAGGTACGATAGACATGGATCTT 4013
Qy    821 IleThrThrGlyValSerAlaSerGluArgIleArgArgAlaAsnLeuLeuAlaAlaLeu 840
Db   4014 ATCACGACTGGAGTGTCGGCCAGCGAGCGTATTCGTCGGGCCAACTTGCTAGCTGCTCTG 4073
Qy    841 ArgGluLeuIleAlaAspLysIleSerProGlySerSerSerGlyLeuLysThrSerGln 860
Db   4074 CGAGAGCTTATAGCAGATAAAATTTCACCTGGCAGCTCCTCTGGCTTGAAGACCAGTCAG 4133
Qy    861 LeuLeuGluAspIleArgSerGlnSerSerValAspValSerLeuGlnAspIleLysAsn 880
Db   4134 CTTCTTGAGGATATCCGGAGCCAAAGCAGTGTGGACGTTAGTTTGCAGGATATTAAAAAT 4193
Qy    881 AlaLeuGlySerLeuGlnGlyGluGlyPheLeuThrValHisGlyAspIleValLysArg 900
Db   4194 GCTCTGGGTAGCCTCCAAGGAGAAGGCTTTCTTACTGTCCATGGTGACATAGTCAAGAGA 4253
Qy    901 Val 901
Db   4254 GTT 4256
```

RESULT 3
US-10-425-114-28735
; Sequence 28735, Application US/10425114
; Publication No. US20040034888A1
; GENERAL INFORMATION:
; APPLICANT: Liu, Jingdong
; APPLICANT: Zhou, Yihua
; APPLICANT: Kovalic, David K.
; APPLICANT: Screen, Steven E
; APPLICANT: Tabaska, Jack E
; APPLICANT: Cao, Yongwei
; TITLE OF INVENTION: Nucleic Acid Molecules and Other Molecules Associated With
; TITLE OF INVENTION: Plants and Uses Thereof for Plant Improvement
; FILE REFERENCE: 38-21(53313)B
; CURRENT APPLICATION NUMBER: US/10/425,114
; CURRENT FILING DATE: 2003-04-28
; NUMBER OF SEQ ID NOS: 73128
; SEQ ID NO 28735
; LENGTH: 2841
; TYPE: DNA
; ORGANISM: Zea mays
; FEATURE:
; OTHER INFORMATION: Clone ID: LIB4763-010-C6_FLI
US-10-425-114-28735

Alignment Scores:

Pred. No.:	8.18e-209	Length:	2841
Score:	2550.00	Matches:	514
Percent Similarity:	73.24%	Conservative:	143
Best Local Similarity:	57.30%	Mismatches:	176
Query Match:	55.33%	Indels:	64
DB:	7	Gaps:	17

```
US-10-768-511-8 (1-901) x US-10-425-114-28735 (1-2841)

Qy     20 SerGluGlyValSerThrProLeuProGlnValThrSerProSerPheAspAsnAlaAla 39
Db    100 TCTGACGGCGGCGGCAGCTCCCCTCCTCCCGCATCGTCGCCC---TATGGTCGGCCGTCA 156

Qy     40 SerProValAlaGlyArgArgAlaValArgGlnThrProThrSerAlaValArgArgArg 59
Db    157 AGCCCACTCGCG---------GTTACTAACTCCTCCCCATCTCAGCCCACACGCCGCCCC 207

Qy     60 GlyArgGluThrAspSerAlaArgArgArgArgSerArgSerArgSerLeuGlyAsnSer 79
Db    208 GGC---------------GGTCGCCGTCGCCGTGGTTCC--------------------- 231

Qy     80 ValTyrSerSerProTyrAspAlaGlyThrProGlyThrProGlyThrProValAlaThr 99
Db    232 ------GCTAGTCCTTAT---GCTTCGTCCCCTTCTCTCGGGGGG--------------- 267

Qy    100 ProValTyrAlaThrProValGlyThrProMetGlyThrProSerPheHisArgGlyThr 119
Db    268 ------TTCGAGACGCCG--------------------CCGCACCCTGGCCGCCGCACG 300

Qy    120 ProGlnTyrLysGlnArgSerGluLeuGlySerGlnGlyLysProLeuHisArgArgArg 139
Db    301 CCG---------------TCCGGAGCTGGGGCT---GGCGCCCCTCGGCAGCCGCGTCAG 342

Qy    140 ArgSerGlnSerArgGluProGlyHisArgSerProSerArgGluProSerAlaAspGly 159
Db    343 AACTCGACTGGCCGGTTCCCG------CCGACGCCCTCCACTCCAATGTCCACCGATGAC 396

Qy    160 ---ArgProSerGluSerAlaGluProAspAspThrLeuGlyGly--------------- 173
Db    397 GTCCCTCCATCCTCCGAAGCTGGGGACGACGAGACCGACGGCGGCGGCGGCGTCGACGCC 456

Qy    174 GluTyrAlaTyrValTrpGlyThrAsnValAsnIleProAspValLeuArgAlaIleArg 193
Db    457 ACCCCGGTCTTCGTCTGGGGCACCAACATAAGCGTGCAGGACGTCAACGCCGCCATTCTC 516

Qy    194 ArgPheLeuHisAsnTyrArgSerSerAlaHisAspLeu--------------------- 206
Db    517 CGGTTCTTGCGCCACTTCCGG---GACCCGCGCGACGCTGGCCGCGTCGACCCGGTCATG 573

Qy    207 ---AsnSerLysTyrIleGlnIleIleGluGluThrValGluArgGlu---GluAspThr 224
Db    574 GACGAGGGCAAGTACATGCGCGCCATCCACCGCATCCTCGAGCTCGAGGGCGGGGAGTCG 633

Qy    225 LeuAsnIleAspMetSerAspIleTyrAspHisAspProAspLeuTyrAlaLysIleVal 244
Db    634 CTCGACGTCGATGCGCACGACGTGTTTGACCACGACCCAGACCTCTACAGCAAGATGGTT 693

Qy    245 ArgTyrProLeuAspIleIleProLeuLeuAspThrGluCysGlnGluValAlaThrSer 264
Db    694 CGCTATCCGCTCGAGGTGCTCGCCATCTTCGACATCGTGCTCATGGACCTCGTCGCGCGC 753

Qy    265 LeuLeuProThrPheGluLysHisIleGluAlaArgProPheAsnLeuLysAlaSerVal 284
Db    754 ATCGAGCCGCTCTTCGAGAAGCACATCCAGACCAGGATCTACAACCTCAAGTCGTCCATT 813

Qy    285 HisMetArgGluLeuAsnProSerAspIleAspLysLeuValSerValLysGlyMetVal 304
Db    814 TGCTTGAGGAATCTCAACCCATCTGATATTGAGAAGATGGTATCCATCAAGGGTATGATA 873

Qy    305 IleArgCysSerSerIleIleProGluIleLysGlyAlaPhePheLysCysLeuValCys 324
Db    874 ATTAGATGCAGCTCGGTCATACCGGAGCTCAAGGAGGCTGTGTTCCGCTGCCTGGTTTGT 933

Qy    325 GlyHisSerProProLeuValThrValValLysGlyArgValGluGluProThrArgCys 344
Db    934 GGTTTCTACTCAGAGCCCGTCATGGTTGATAGAGGAAGAGTAACTGAACCACACATTTGT 993

Qy    345 GluLysProGluCysAlaAlaArgAsnAlaMetSerLeuIleHisAsnArgCysThrPhe 364
Db    994 CAGAAAGAACAATGTAAAGCCACAAATTCTATGACCCTAGTGCACAACAGATGCAGATTT 1053

Qy    365 AlaAsnLysGlnIleValArgLeuGlnGluThrProAspAlaIleProGluGlyGluThr 384
Db   1054 TCAGACAAGCAGATCATAAAGTTGCAGGAAACACCAGACGAGATACCAGAAGGTGGCACT 1113

Qy    385 ProHisThrValSerMetCysLeuTyrAsnThrMetValAspAlaValLysProGlyAsp 404
Db   1114 CCACATACAGTTAGTGTCTTGATGCATGATAAGCTTGTTGATGCTGGAAAGCCTGGAGAT 1173

Qy    405 ArgIleGluValThrGlyValPheLysAlaMetAlaValArgValGlyProAsnGlnArg 424
Db   1174 AGGGTTGAGATAACTGGAATATACAGAGCTATGAGTATTCGAATTGGACCAACTCAAAGG 1233

Qy    425 ThrLeuArgAlaLeuTyrLysThrTyrIleAspCysValHisValLysLysSerAspArg 444
Db   1234 ACAGTGAAGTCTATATTCAAGACATATATTGATTGCCTTCACATAAAGAAGACAGACAAG 1293

Qy    445 GlyArgLeuGlnThrGluAspProMetGluMetAspLysGluAsnAspMetTyrAlaGly 464
Db   1294 TCTAGGCTTCATGTGGAGGACACCATGGATATTGATAATTCTAAC--------------- 1338

Qy    465 TyrHisGluSerAspThrSerGluAlaAlaAsnGluAlaLysIleGlnLysLeuLysGlu 484
Db   1339 ---GCTAGCAAATCTACTGAAGAGGATTTTCTTAGTGATAAGGTTGAGAAACTAAAAGAG 1395

Qy    485 LeuSerLysLeuProGlyIleTyrAspArgLeuSerArgSerLeuAlaProSerIleTrp 504
Db   1396 CTTTCGAAGTTGCCTGATATCTATGAAAGATTGACTAGATCATTAGCTCCAAACATATGG 1455

Qy    505 GluLeuGluAspIleLysLysGlyLeuLeuCysGlnLeuPheGlyGlyLysAlaLysLys 524
Db   1456 GAGTTGGATGATGTCAAAAGAGGTCTCCTTTGCCAGCTTTTCGGCGGTAATCCCTTGAAG 1515

Qy    525 IleProSerGlyAlaSerPheArgGlyAspIleAsnValLeuLeuValGlyAspProGly 544
Db   1516 CTTCCTTCTGGAGCTAGTTTCCGGGGTGACATCAATATTTTACTTGTGGGGGACCCTGGA 1575

Qy    545 ThrSerLysSerGlnLeuLeuGlnTyrValHisLysIleAlaProArgGlyIleTyrThr 564
Db   1576 ACAAGTAAATCCCAGCTTCTCCAGTACATGCATAAACTGTCTCCTCGTGGTATCTATACG 1635

Qy    565 SerGlyArgGlySerSerAlaValGlyLeuThrAlaTyrValThrLysAspProGluThr 584
Db   1636 AGTGGTAGAGGAAGTTCTGCTGTTGGTCTTACTGCTTATGTTACCAAAGACCCTGAGACT 1695

Qy    585 ArgGluThrValLeuGluSerGlyAlaLeuValLeuSerAspArgGlyIleCysCysIle 604
Db   1696 GGCGAAACTGTTCTAGAAAGTGGAGCACTTGTTTTGAGTGACAAAGGTGTTTGTTGCATA 1755

Qy    605 AspGluPheAspLysMetSerAspAsnAlaArgSerMetLeuHisGluValMetGluGln 624
Db   1756 GATGAGTTTGATAAGATGTCTGATAATGCCCGAAGCATGTTACACGAGGTGATGGAACAG 1815

Qy    625 GlnThrValSerValAlaLysGlyGlyIleIleAlaSerLeuAsnAlaArgThrSerVal 644
Db   1816 CAGACAGTATCCATTGCGAAGGCTGGAATAATTGCATCTTTAAACGCTAGGACATCTGTC 1875

Qy    645 LeuAlaCysAlaAsnProSerGlySerArgTyrAsnAlaArgLeuSerValIleAspAsn 664
Db   1876 CTGGCATGTGCCAATCCTACTGAATCACGTTACAATCCAAGGCTCTCTGTAATTGACAAC 1935

Qy    665 IleGlnLeuProProThrLeuLeuSerArgPheAspLeuIleTyrLeuMetLeuAspLys 684
Db   1936 ATCCACTTAGCGCCAACGCTACTTTCAAGATTCGACCTGATTTATCTTATCTTGGACAAG 1995

Qy    685 ProAspGluGlnAsnAspArgArgLeuAlaArgHisLeuValAlaLeuHisTyrGluAsn 704
Db   1996 GCGGATGAGCAAACTGATAGGCGCCTGGCAAAGCATATTGTTTCGTTGCATTTTGAGAAT 2055

Qy    705 TyrGluValSerLysGlnAspAlaLeuAspLeuGlnThrLeuThrAlaTyrIleThrTyr 724
Db   2056 CCAAATTTAGAGGAGCTCGAGGTCTTGGACTTGCAGACACTAGTTTCCTACATAAGCTAT 2115
```

```
Qy        725 AlaArgGlnHisValHisProThrLeuSerAspGluAlaAlaGluAspLeuIleAsnGly 744
              ||||||::::::::   |||   |||||||||||||||||||||:::|||       |||
Db       2116 GCAAGGAAGTATATTCAGCCACAGTTATCTGATGAAGCTGCAGAAGAGTTAACTCGTGGC 2175

Qy        745 TyrValGluMetArgGlnLysGlyAsnPheProGlySerSerLysLysValIleThrAla 764
              |||||||||||||||::::::||||||   |||||||||   ||||||||||||||||||
Db       2176 TATGTCGAGATGAGAAAAAGAGGGAATAGCCCTGGGAGCAGAAAGAAGGTCATAACAGCA 2235

Qy        765 ThrProArgGlnLeuGluSerMetIleArgIleSerGluAlaLeuAlaArgMetArgPhe 784
              |||   ||||||:::||||||:::||||||:::|||||||||  |||||||||||||||
Db       2236 ACCGCTAGACAAATAGAGAGTTTGATCCGTCTCAGCGAAGCATTAGCCCGAATGCGGTTC 2295

Qy        785 SerGluValValGluLysValAspAlaAlaGluAlaValArgLeuLeuAspValAlaLeu 804
              ||||||||||||||         |||       ||||||    |||||||||:::||||||:::
Db       2296 TCTGAAGTGGTCGAGGTGCGGGATGTTGTGGAGGCATTCAGGCTTCTTGAAGTCGCCATG 2355

Qy        805 GlnGlnSerAlaThrAspHisAlaThrGlyThrIleAspMetAspLeuIleThrThrGly 824
              ||||||||||||||||||||||||||||||||||||||||||||||||||     ||||||
Db       2356 CAGCAGTCTGCGACGGATCATGCAACTGGTACGATTGATATGGATCTGATCATGACGGGG 2415

Qy        825 ValSerAlaSerGluArgIleArgArgAlaAsnLeuLeuAlaAlaLeuArgGluLeuIle 844
              :::||||||||||||||||     ||||||    ||||||:::||||||  |||   ||||||
Db       2416 ATATCCGCAAGCGAAAGGCAGAGGCGGGAGAACCTCGTTGCCGCAACCCGTAACCTGATT 2475

Qy        845 AlaAspLysIleSerProGlySerSerSerGlyLeuLysThrSerGlnLeuLeuGluAsp 864
              |||:::|||:::      |||       |||    ::::::     :::|||||||||:::
Db       2476 GCGGAGAAAATGCAGCTTGGAGGCCCCTCG---ATGCGCATGATTGAGTTGCTGGAGGAA 2532

Qy        865 IleArgSerGlnSerSerValAspValSerLeuGlnAspIleLysAsnAlaLeuGlySer 884
              :::|||   |||||||||:::::::::   :::    :::::::::   |||||||||:::
Db       2533 CTGAGGAAGCAGAGCTCAATGGAAATTCATATGCACGAACTCCGCGGTGCTCTTGGCACC 2592

Qy        885 LeuGlnGlyGluGlyPheLeuThrValHisGlyAspIleValLysArgVal 901
              |||      ||||||    :::    :::|||||||||    |||:::||||||
Db       2593 CTGATGACTGAAGGCGCGGTGGTTATCCATGGAGACAACGTGAGGAGAGTT 2643
```

RESULT 4
US-10-437-963-44639
; Sequence 44639, Application US/10437963
; Publication No. US20040123343A1
; GENERAL INFORMATION:
; APPLICANT: La Rosa, Thomas J.
; APPLICANT: Kovalic, David K.
; APPLICANT: Zhou, Yihua
; APPLICANT: Cao, Yongwei
; APPLICANT: Wu, Wei
; APPLICANT: Boukharov, Andrey A.
; APPLICANT: Barbazuk, Brad
; APPLICANT: Li, Ping
; TITLE OF INVENTION: Rice Nucleic Acid Molecules and Other Molecules Associated With
; TITLE OF INVENTION: Plants and Uses Thereof for Plant Improvement
; FILE REFERENCE: 38-21(53221)B
; CURRENT APPLICATION NUMBER: US/10/437,963
; CURRENT FILING DATE: 2003-05-14
; NUMBER OF SEQ ID NOS: 204966
; SEQ ID NO 44639
;   LENGTH: 3054
;   TYPE: DNA
;   ORGANISM: Oryza sativa
;   FEATURE:
;   OTHER INFORMATION: Clone ID: PAT_MRT4530_47680C.1
US-10-437-963-44639

Alignment Scores:

Pred. No.:	7.06e-203	Length:	3054
Score:	2481.50	Matches:	512
Percent Similarity:	71.96%	Conservative:	132
Best Local Similarity:	57.21%	Mismatches:	180
Query Match:	53.84%	Indels:	73
DB:	7	Gaps:	15

US-10-768-511-8 (1-901) x US-10-437-963-44639 (1-3054)

```
Qy         24 SerThrProLeuProGlnValThrSerProSerPheAspAsnAlaAlaSerProValAla 43
              ::::::|||   |||   |||:::||||||                  :::|||||||:::
Db         37 AATTCCCCTCCTCCCTCCGTGTCGTCGCCGGATGTG---CGACCGTCGAGCCCTCTCCCG 93

Qy         44 GlyArgArgAlaValArgGlnThrProThrSerAlaValArgArgArgGlyArgGluThr 63
                          |||         :::|||     :::    ||||||   |||
Db         94 ---------GCGACCAACTCCTCCCCGCCTCAGTCGGGGCGGCGCGGCGGG--------- 135

Qy         64 AspSerAlaArgArgArgArgSerArgSerArgSerLeuGlyAsnSerValTyrSerSer 83
                       ||||||||||||                               :::|||
Db        136 ------GGGCGGCGCCGGCGTGGTTCC-------------------------GCCAGC 162

Qy         84 ProTyrAspAlaGlyThrProGlyThrProGlyThrProValAlaThrProValTyrAla 103
              ||||||   :::    :::|||             |||                  :::
Db        163 CCGTACCCGTCG---TCCCCGTCCCTCGGCGGG--------------------TTCGAG 198

Qy        104 ThrProValGlyThrProMetGlyThrProSerPheHisArgGlyThrProGlnTyrLys 123
              ||||||                   |||||||||          |||           :::
Db        199 ACGCCGCCGCACCCGGGCCGCCGCACGCCCTCC------GGCGGCGCGGCGGCGCGGCAG 252

Qy        124 GlnArgSerGluLeuGlySerGlnGlyLysPro---------------------LeuHis 136
              ||||||   ||||||                                      :::|||
Db        253 CAGCGGCA-GAACTGGACGGGGGGGCGGTTCCCGCCGACCCCGTCCACCCCCATGTCCAC 311

Qy        137 ArgArgArgArgSerGlnSerArgGluProGlyHisArgSerProSerArgGluProSer 156
              |||||||||            |||   ||||||   |||:::|||   |||        :::
Db        312 CGACGACGTCC-CGCTCTCCTCGGAGGCCGGGGACGAGGACACCCCCGAGACCGACGGCG 370

Qy        157 Ala--AspGlyArgProSerGluSerAlaGluProAspAspThrLeuGlyGlyGluTyrA 176
              |||    |||                   |||   |||
Db        371 GCGGCGGCGGCGGCGCGGGCGCCGACGCCACGCCG-----------------------G 406

Qy        176 laTyrValTrpGlyThrAsnValAsnIleProAspValLeuArgAlaIleArgArgPheL 196
                 :::|||||||||||||||||:::::::::      ||||||      ||||||   |||||||
Db        407 TGTTCGTCTGGGGCACCAACATCAGCGTGCAGGACGTCAACGCCGCCATCCTCCGGTTCC 466

Qy        196 euHisAsnTyrArgSerSerAlaHisAspLeu---------------------AsnS 208
              ||    ::::::|||              |||
Db        467 TCCGCCACTTCCGC---GACCCGCGCGACGCCGGCCGCGTCGACCCGGTCATGGACGAGG 523

Qy        208 erLysTyrIleGlnIleIleGluGluThrValGluArgGlu---GluAspThrLeuAsnI 227
                 ||||||::::::   |||         :::|||   |||      ::::::|||::::
Db        524 GCAAGTACATGCGCGCCATCCACCGCATCCTCGAGCTCGAGGGCGGGGAGTCGCTCGACG 583

Qy        227 leAspMetSerAspIleTyrAspHisAspProAspLeuTyrAlaLysIleValArgTyrP 247
              :::::         |||::::::||||||||||||||||||||   |||:::|||||||||
Db        584 TCAACGCCCACGACGTGTTCGACCACGACCCGGACCTCTACGGCAAGATGGTCAGGTACC 643

Qy        247 roLeuAspIleIleProLeuLeuAspThrGluCysGlnGluValAlaThrSerLeuLeuP 267
              |||||:::::::::   :::   |||                 ::::::         :::   |
Db        644 CGCTCGAGGTGCTCGCCATCTTTGACATCGTGCTCATGGACCTCGTCGCGCGCATCGAGC 703

Qy        267 roThrPheGluLysHisIleGluAlaArgProPheAsnLeuLysAlaSerValHisMetA 287
              ||      |||||||||||||||:::      |||      :::||||||||:::||||||   :::|
Db        704 CCCTCTTCGAGAAGCACATCCAGACCAGGATCTACAACCTCAAGTCCTCGGTTTGCTTGA 763

Qy        287 rgGluLeuAsnProSerAspIleAspLysLeuValSerValLysGlyMetValIleArgC 307
              ||   |||||||||||||||||:::|||:::||||||:::|||||||||:::|||||||
Db        764 GGAATCTCAACCCTTCTGATATCGAGAAGATGGTGTCCATCAAGGGTATGATAATTCGGT 823

Qy        307 ysSerSerIleIleProGluIleLysGlyAlaPhePheLysCysLeuValCysGlyHisS 327
              |||||||:::|||||||||:::|||   |||   |||:::|||||||||||||||
Db        824 GCAGCTCGGTGATTCCAGAGCTCAAGGAGGCTGTGTTCCGCTGCTTGGTTTGCGGGTTCT 883

Qy        327 erProProLeuValThrValValLysGlyArgValGluGluProThrArgCysGluLysP 347
                          |||   |||   :::|||||||||   ||||||         |||:::|||
Db        884 ACTCTGAGCCTGTAATGGTTGATCGAGGGAGGGTTACTGAGCCGCACATCTGTCAGAAAG 943

Qy        347 roGluCysAlaAlaArgAsnAlaMetSerLeuIleHisAsnArgCysThrPheAlaAsnL 367
                 :::|||   |||   |||:::|||:::|||:::||||||||||||   ||||||:::|
Db        944 AACAATGTAAAGCTACAAACTCTATGACTCTTGTGCATAACCGATGCAGGTTTGCAGATA 1003
```

```
Qy      367 ysGlnIleValArgLeuGlnGluThrProAspAlaIleProGluGlyGluThrProHisT 387
Db     1004 AGCAGATCATAAAGTTGCAGGAAACACCAGATGAGATACCAGAAGGTGGCACTCCACATA 1063

Qy      387 hrValSerMetCysLeuTyrAsnThrMetValAspAlaValLysProGlyAspArgIleG 407
Db     1064 CTGTTAGTGTCTTGATGCACGATAAGCTTGTAGATGCTGGAAAGCCTGGAGACAGGGTTG 1123

Qy      407 luValThrGlyValPheLysAlaMetAlaValArgValGlyProAsnGlnArgThrLeuA 427
Db     1124 AGATAACTGGGATATACAGAGCCATGAGTATTAGAGTTGGCCCAACTCAGAGGACAGTGA 1183

Qy      427 rgAlaLeuTyrLysThrTyrIleAspCysValHisValLysLysSerAspArgGlyArgL 447
Db     1184 AGTCGATTTTC--------------------------AAGAAGACAGACAAGTCTAGAC 1216

Qy      447 euGlnThrGluAspProMetGluMetAspLysGluAsnAspMetTyrAlaGlyTyrHisG 467
Db     1217 TTCATGTTGAGGACTCCATGGAAACTGATAACCCCAAT-----------------GCTA 1258

Qy      467 luSerAspThrSerGluAlaAlaAsnGluAlaLysIleGlnLysLeuLysGluLeuSerL 487
Db     1259 ACAAGACAACTGAAGATGATTTTCTCAGAGATAAGGTTGAGAAATTAAAAGAGTTGTCAA 1318

Qy      487 ysLeuProGlyIleTyrAspArgLeuSerArgSerLeuAlaProSerIleTrpGluLeuG 507
Db     1319 AGTTGCCAGATATATATGACAGATTAACTAGGTCATTGGCTCCAAACATATGGGAGCTGG 1378

Qy      507 luAspIleLysLysGlyLeuLeuCysGlnLeuPheGlyGlyLysAlaLysLysIleProS 527
Db     1379 ACGACGTTAAAAGGGGCCTCCTTTGCCAGCTTTTTGGTGGAAATGCTTTGAGGCTTCCTT 1438

Qy      527 erGlyAlaSerPheArgGlyAspIleAsnValLeuLeuValGlyAspProGlyThrSerL 547
Db     1439 CTGGAGCTAGTTTCCGAGGCGACATCAATATTTTGCTTGTTGGTGATCCTGGAACAAGTA 1498

Qy      547 ysSerGlnLeuLeuGlnTyrValHisLysIleAlaProArgGlyIleTyrThrSerGlyA 567
Db     1499 AATCCCAGCTTCTCCAATACATGCACAAACTGTCTCCTCGTGGCATTTATACAAGTGGCA 1558

Qy      567 rgGlySerSerAlaValGlyLeuThrAlaTyrValThrLysAspProGluThrArgGluT 587
Db     1559 GAGGAAGTTCAGCTGTTGGCCTTACTGCATACGTTACCAAGGATCCTGAAACTGGTGAAA 1618

Qy      587 hrValLeuGluSerGlyAlaLeuValLeuSerAspArgGlyIleCysCysIleAspGluP 607
Db     1619 CTGTTCTTGAGAGTGGAGCGCTTGTTTTGAGTGACAAAGGTGTTTGTTGTATTGATGAAT 1678

Qy      607 heAspLysMetSerAspAsnAlaArgSerMetLeuHisGluValMetGluGlnGlnThrV 627
Db     1679 TTGATAAGATGTCTGATAATGCTCGAAGCATGTTACATGAGGTGATGGAACAACAGACTG 1738

Qy      627 alSerValAlaLysGlyGlyIleIleAlaSerLeuAsnAlaArgThrSerValLeuAlaC 647
Db     1739 TCTCCATTGCCAAGGCTGGAATAATTGCATCTTTAAATGCCAGAACATCAGTTCTAGCAT 1798

Qy      647 ysAlaAsnProSerGlySerArgTyrAsnAlaArgLeuSerValIleAspAsnIleGlnL 667
Db     1799 GTGCAAATCCTACTGAATCACGTTATAATCCAAGGCTCTCTGTGATTGACAATATCCATC 1858

Qy      667 euProProThrLeuLeuSerArgPheAspLeuIleTyrLeuMetLeuAspLysProAspG 687
Db     1859 TTCCTCCAACACTGCTTTCTAGGTTTGACCTCATTTATCTGATATTGGACAAGGCAGATG 1918

Qy      687 luGlnAsnAspArgArgLeuAlaArgHisLeuValAlaLeuHisTyrGluAsnTyrGluV 707
Db     1919 AGCAAACTGATAGACGCCTGGCTAAGCATATTGTTTCGTTGCATTTTGAGAATCCAAACA 1978

Qy      707 alSerLysGlnAspAlaLeuAspLeuGlnThrLeuThrAlaTyrIleThrTyrAlaArgG 727
Db     1979 TAGAGGAGCTCGAGGTCTTGGATTTGCCAACACTTGTAGCCTACATAAGTTATGCAAGGA 2038

Qy      727 lnHisValHisProThrLeuSerAspGluAlaAlaGluAspLeuIleAsnGlyTyrValG 747
Db     2039 AGCATATACAACCACAGTTATCTGATGAAGCTGCAGAAGAATTGACTCGCGGCTATGTTG 2098

Qy      747 luMetArgGlnLysGlyAsnPheProGlySerSerLysLysValIleThrAlaThrProA 767
Db     2099 AGATGAGGAAAAGAGGAAACAGCCCTGGTAGCAGAAAGAAGGTCATAACTGCGACAGCTC 2158

Qy      767 rgGlnLeuGluSerMetIleArgIleSerGluAlaLeuAlaArgMetArgPheSerGluV 787
Db     2159 GACAAATTGAGAGCTTGATTCGGCTCAGTGAAGCACTGGCCCGAATGCGATTCTCTGAAA 2218

Qy      787 alValGluLysValAspAlaAlaGluAlaValArgLeuLeuAspValAlaLeuGlnGlnS 807
Db     2219 TGGTTGAAGTACAAGATGTTGTAGAGGCCTTCAGGCTTCTCGAAGTTGCCATGCAGCAAT 2278

Qy      807 erAlaThrAspHisAlaThrGlyThrIleAspMetAspLeuIleThrThrGlyValSerA 827
Db     2279 CGGCAACTGATCATGCCACTGGTACAATCGATATGGATCTTATCATGACTGGAATATCTG 2338

Qy      827 laSerGluArgIleArgArgAlaAsnLeuLeuAlaAlaLeuArgGluLeuIleAlaAspL 847
Db     2339 CGAGCGAAAGGCAGAGGCGGGACAATCTCGTTGCAGCAACCCGCAACCTTGTGATGGAGA 2398

Qy      847 ysIleSerProGlySerSerSerGlyLeuLysThrSerGlnLeuLeuGluAspIleArgS 867
Db     2399 AAATGCAGCTTGGAGGGCCCTCA---GTGCGAATGATTGAGTTGCTGGAAGAAATTAGGA 2455

Qy      867 erGlnSerSerValAspValSerLeuGlnAspIleLysAsnAlaLeuGlySerLeuGlnG 887
Db     2456 AGCAGAGCTCTATGGAAGTTCATCTGCATGATCTTCGCGGTGCTCTTGGCACTCTGATGA 2515

Qy      887 lyGluGlyPheLeuThrValHisGlyAspIleValLysArg 900
Db     2516 CAGAGGGTGCCGTAGTCATCCATGGAGACAGCGTCAAGAGG 2556
```

RESULT 5
US-10-425-115-98898
; Sequence 98898, Application US/10425115
; Publication No. US20040214272A1
; GENERAL INFORMATION:
; APPLICANT: La Rosa, Thomas J.
; APPLICANT: Kovalic, David K.
; APPLICANT: Zhou, Yihua
; APPLICANT: Cao, Yongwei
; TITLE OF INVENTION: Nucleic Acid Molecules and Other Molecules Associated With
; TITLE OF INVENTION: Plants
; FILE REFERENCE: 38-21(53222)B
; CURRENT APPLICATION NUMBER: US/10/425,115
; CURRENT FILING DATE: 2003-04-28
; NUMBER OF SEQ ID NOS: 369326
; SEQ ID NO 98898
; LENGTH: 2186
; TYPE: DNA
; ORGANISM: Zea mays
; FEATURE:
; NAME/KEY: unsure
; LOCATION: (1)..(2186)
; OTHER INFORMATION: unsure at all n locations
; FEATURE:
; OTHER INFORMATION: Clone ID: MRT4577_21708C.1
US-10-425-115-98898

Alignment Scores:

Pred. No.:	1.67e-158	Length:	2186
Score:	1963.50	Matches:	381
Percent Similarity:	81.79%	Conservative:	86
Best Local Similarity:	66.73%	Mismatches:	81
Query Match:	42.60%	Indels:	24
DB:	8	Gaps:	3

US-10-768-511-8 (1-901) x US-10-425-115-98898 (1-2186)

```
Qy      331 ValThrValValLysGlyArgValGluGluProThrArgCysGluLysProGluCysAla 350
```

```
Db         11 GTCATGGTTGATAGAGGAAGAGTAACTGAACCGCACATTTGTCAGAAAGAACAATGTAAA 70

Qy        351 AlaArgAsnAlaMetSerLeuIleHisAsnArgCysThrPheAlaAsnLysGlnIleVal 370

Db         71 GCCACGAATTCTATGACCCTAGTGCACAACCGATGCAGATTTTCAGACAAGCAGATCATA 130

Qy        371 ArgLeuGlnGluThrProAspAlaIleProGluGlyGluThrProHisThrValSerMet 390

Db        131 AAGTTGCAGGAAACACCAGACGAGATACCAGAAGGTGGCACTCCACATACAGTTAGTGTC 190

Qy        391 CysLeuTyrAsnThrMetValAspAlaValLysProGlyAspArgIleGluValThrGly 410

Db        191 TTGATGCATGATAAGCTTGTTGATGCTGGAAAGCCTGGAGATAGGGTTGAGATAACTGGA 250

Qy        411 ValPheLysAlaMetAlaValArgValGlyProAsnGlnArgThrLeuArgAlaLeuTyr 430

Db        251 ATATACAGAGCTATGAGTATTCGAATTGGACCAAC------------------------ 285

Qy        431 LysThrTyrIleAspCysValHisValLysLysSerAspArgGlyArgLeuGlnThrGlu 450

Db        286 -----------------------ATAAAGAAGACAGACAAGTCTAGGCTTCATGTGGAG 321

Qy        451 AspProMetGluMetAspLysGluAsnAspMetTyrAlaGlyTyrHisGluSerAspThr 470

Db        322 GACACCATGGATATTGATAATTCTAAC-----------------GCTAGCAAATCTACT 363

Qy        471 SerGluAlaAlaAsnGluAlaLysIleGlnLysLeuLysGluLeuSerLysLeuProGly 490

Db        364 GAAGAGGATTTTCTTAGTGATAAGGTTGAGAAACTAAAAGAGCTTTCGAAGTTGCCTGAT 423

Qy        491 IleTyrAspArgLeuSerArgSerLeuAlaProSerIleTrpGluLeuGluAspIleLys 510

Db        424 ATCTATGAAAGATTGACTAGATCATTAGCTCCAAACATATGGGAGTTGGATGATGTCAAA 483

Qy        511 LysGlyLeuLeuCysGlnLeuPheGlyGlyLysAlaLysLysIleProSerGlyAlaSer 530

Db        484 AGAGGTCTCCTTTGCCAGCTTTTCGGCGGCAATCCCTTGAAGCTTCCTTCTGGAGCTAGT 543

Qy        531 PheArgGlyAspIleAsnValLeuLeuValGlyAspProGlyThrSerLysSerGlnLeu 550

Db        544 TTCCGGGGTGACATCAATATTTTACTTGTGGGGGACCCTGGAACAAGTAAATCCCAGCTT 603

Qy        551 LeuGlnTyrValHisLysIleAlaProArgGlyIleTyrThrSerGlyArgGlySerSer 570

Db        604 CTCCAGTACATGCATAAACTGTCTCCTCGTGGTATCTATACGAGTGGTAGAGGAAGTTCT 663

Qy        571 AlaValGlyLeuThrAlaTyrValThrLysAspProGluThrArgGluThrValLeuGlu 590

Db        664 GCTGTTGGTCTTACTGCTTATGTTACCAAAGACCCTGAGACTGGCGAAACTGTTCTAGAA 723

Qy        591 SerGlyAlaLeuValLeuSerAspArgGlyIleCysCysIleAspGluPheAspLysMet 610

Db        724 AGTGGAGCACTTGTTTTGAGTGACAAAGGTGTTTGTTGCATAGATGAGTTTGATAAGATG 783

Qy        611 SerAspAsnAlaArgSerMetLeuHisGluValMetGluGlnGlnThrValSerValAla 630

Db        784 TCTGATAATGCCCGAAGCATGTTACACGAGGTGATGGAACAGCAGACAGTATCCATTGCG 843

Qy        631 LysGlyGlyIleIleAlaSerLeuAsnAlaArgThrSerValLeuAlaCysAlaAsnPro 650

Db        844 AAGGCTGGAATAATTGCATCTTTAAACGCTAGGACATCTGTCCTGGCATGTGCCAATCCT 903

Qy        651 SerGlySerArgTyrAsnAlaArgLeuSerValIleAspAsnIleGlnLeuProProThr 670

Db        904 ACTGAATCACGTTACAATCCAAGGCTCTCTGTAATTGACAACATCCACTTAGCGCCAACG 963

Qy        671 LeuLeuSerArgPheAspLeuIleTyrLeuMetLeuAspLysProAspGluGlnAsnAsp 690

Db        964 CTACTTTCAAGATTCGACCTGATTTATCTTATCTTGGACAAGGCGGATGAGCAAACTGAT 1023

Qy        691 ArgArgLeuAlaArgHisLeuValAlaLeuHisTyrGluAsnTyrGluValSerLysGln 710

Db       1024 AGGCGCCTGGCAAAGCATATTGTTTCGTTGCATTTTGAGAATCCAAATTTAGAGGAGCTC 1083

Qy        711 AspAlaLeuAspLeuGlnThrLeuThrAlaTyrIleThrTyrAlaArgGlnHisValHis 730

Db       1084 GAGGTCTTGGACTTGCAGACACTAGTTTCCTACATAAGCTATGCAAGGAAGTATATTCAG 1143

Qy        731 ProThrLeuSerAspGluAlaAlaGluAspLeuIleAsnGlyTyrValGluMetArgGln 750

Db       1144 CCACAGTTATCTGATGAAGCTGCAGAAGAGTTAACTCGTGGCTATGTCGAGATGAGAAAA 1203

Qy        751 LysGlyAsnPheProGlySerSerLysLysValIleThrAlaThrProArgGlnLeuGlu 770

Db       1204 AGAGGGAATAGCCCTGGGAGCAGAAAGAAGGTCATAACAGCAACCGCTAGACAAATAGAG 1263

Qy        771 SerMetIleArgIleSerGluAlaLeuAlaArgMetArgPheSerGluValValGluLys 790

Db       1264 AGTTTGATCCGTCTCAGCGAAGCATTAGCCCGAATGCGGTTCTCTGAAGTGGTCGAGGTG 1323

Qy        791 ValAspAlaAlaGluAlaValArgLeuLeuAspValAlaLeuGlnGlnSerAlaThrAsp 810

Db       1324 CGGGATGTTGTGGAGGCATTCAGGCTTCTTGAAGTCGCCATGCAGCAGTCTGCGACGGAT 1383

Qy        811 HisAlaThrGlyThrIleAspMetAspLeuIleThrThrGlyValSerAlaSerGluArg 830

Db       1384 CATGCAACTGGGACGATTGATATGGATCTGATCATGACGGGGATATCCGCAAGCGAAAGG 1443

Qy        831 IleArgArgAlaAsnLeuLeuAlaAlaLeuArgGluLeuIleAlaAspLysIleSerPro 850

Db       1444 CAGAGGCGGGAGAACCTCGTTGCCGCAACCCGTAACCTGATTGCGGAGAAAATGCAGCTT 1503

Qy        851 GlySerSerSerGlyLeuLysThrSerGlnLeuLeuGluAspIleArgSerGlnSerSer 870

Db       1504 GGAGGCCCCTCG---ATGCGCATGATTGAGTTGCTGGAGGAACTGAGGAAGCAGAGCTCA 1560

Qy        871 ValAspValSerLeuGlnAspIleLysAsnAlaLeuGlySerLeuGlnGlyGluGlyPhe 890

Db       1561 ATGGAAATTCATATGCACGAACTCCGCGGTGCTCTTGGCACCCTGATGACTGAAGGCGCG 1620

Qy        891 LeuThrValHisGlyAspIleValLysArgVal 901

Db       1621 GTGGTTATCCATGGAGACAACGTGAGGAGAGTT 1653
```

RESULT 6
US-10-128-714-7558
; Sequence 7558, Application US/10128714
; Publication No. US20030119013A1
; GENERAL INFORMATION:
; APPLICANT: Jiang, Bo
; APPLICANT: Hu, Wenqi
; APPLICANT: Tishkoff, Daniel
; APPLICANT: Zamudio, Carlos
; APPLICANT: Eroshkin, Alexey M
; APPLICANT: Lemieux, Sebastien M
; TITLE OF INVENTION: Identification of Essential Genes in Aspergillus fumigatus and
; TITLE OF INVENTION: Methods of Use
; FILE REFERENCE: 10182-018-999
; CURRENT APPLICATION NUMBER: US/10/128,714
; CURRENT FILING DATE: 2002-04-23
; PRIOR APPLICATION NUMBER: US 60/285,697
; PRIOR FILING DATE: 2001-04-23
; PRIOR APPLICATION NUMBER: US 60/287,066
; PRIOR FILING DATE: 2001-04-27
; PRIOR APPLICATION NUMBER: US 60/295,890
; PRIOR FILING DATE: 2001-06-05
; PRIOR APPLICATION NUMBER: US 60/303,899
; PRIOR FILING DATE: 2001-07-09
; PRIOR APPLICATION NUMBER: US 60/316,362
; PRIOR FILING DATE: 2001-08-31
; NUMBER OF SEQ ID NOS: 8603
; SOFTWARE: PatentIn version 3.1
; SEQ ID NO 7558
;   LENGTH: 3072
;   TYPE: DNA

```
;   ORGANISM: Aspergillus fumigatus
US-10-128-714-7558

Alignment Scores:
Pred. No.:               2.23e-151     Length:         3072
Score:                   1883.00       Matches:        438
Percent Similarity:      57.77%        Conservative:   157
Best Local Similarity:   42.52%        Mismatches:     278
Query Match:             40.85%        Indels:         159
DB:                      5             Gaps:           23

US-10-768-511-8 (1-901) x US-10-128-714-7558 (1-3072)

Qy     12 ValSerSerProTyrProSerGlnSerGluGlyValSerThrProLeuProGlnValThr 31
Db      1 ATGTCTTCTCCAGCTTCTTCTCGCCGTAGGGGC--------------------------- 33

Qy     32 SerProSerPheAspAsnAlaAlaSerProValAlaGlyArgArgAlaValArgGlnThr 51
Db     34 CGCCCAGCCAAGGACACGGCAACCTCTTCTCCCGCGCGGTCAACCCGTTCTCGTCAGCTC 93

Qy     52 ProThrSerAlaValArgArgArgGlyArgGluThrAspSerAlaArgArgArgArgSer 71
Db     94 CAGACCAGTA--GTCCGACTCCAAGGGCGGCAGATGAACAGTCACAGGCGACGCCT-CGT 150

Qy     72 ArgSerArgSerLeuGlyAsnSerVal-----------------------------Tyr 81
Db    151 GCTTCAAGGCGACTGAGAGGCGAAGCGGCAGTTCCCTCGTCATCGCCTATGTTCTTCCAG 210

Qy     82 SerSerProTyrAspAlaGlyThrProGly---ThrProGlyThrProValAlaThrPro 100
Db    211 TCCTCCCCTTCAAAGGCAGACAGCAGTGCGGAAACACCAGATGTTCGAATGGATGAGCCC 270

Qy    101 ValTyrAlaThrProVal-----------------------------GlyThrProMet 110
Db    271 ------AGTTCACCAATGCGAGAATCTTCTACTATGGATGAAGGAGACAGGACACCTAGA 324

Qy    111 Gly---ThrProSerPheHisArgGlyThrPro-------------------------- 120
Db    325 GGAAATGCTCCTACCATGAGAGACTCCTCCCCCATCCGGTATATGTCTAGCTCCAGTCCA 384

Qy    121 ---------------GlnTyrLysGlnArgSerGluLeuGlySerGlnGlyLysProLeu 135
Db    385 ACTCGAGCTCAGAATCGCCAGTCCCGGCGTTCCGACATTCCCAGCAGCAGCAGCGGATTA 444

Qy    136 His-----------------------------------ArgArgArgArgSerGlnSer 143
Db    445 TTCGTCTCGTCCAGACCCAGCATCGAGAGCAACCGTGCCGTGTCCCGCCGTAGCGACCTT 504

Qy    144 ArgGluProGlyHis---ArgSerProSerArgGlu----------------------- 154
Db    505 CATTCTGGTGGTTTTCTTTCTAGCCCGAACCGTCGCCGCAGAGTTTTTGTCGATGCTAAT 564

Qy    155 ------ProSerAlaAspGlyArgProSerGluSerAla---------------GluPro 167
Db    565 GGTATGCCTGCGACCGATGGCGATCCACGTTCCGATGCCACCTTCTCGAATATCCACCCA 624

Qy    168 Asp-----------AspThrLeuGlyGlyGluTyrAla---TyrValTrpGlyThrAsn 182
Db    625 GACACCTCTGAGGCCGAGGCCTTGGGCGGTAGCTCAACCCGTGTGATTTGGGGTACCAAC 684

Qy    183 ValAsnIleProAspValLeuArgAlaIleArgArgPheLeuHisAsn------------ 198
Db    685 ATCTCCATCCAGGATTCCATGTCCGCATTCAAGAACTTTCTCTACAACTTCCAAACAAAA 744

Qy    199 TyrArgSerSerAlaHisAspLeuAsnSerLysTyrIleGlnIleIleGluGluThrVal 218
Db    745 TACCGTCTGTGGGCAGAGGGCGCAACCGAGGATGAGACACGTATAATGGGTGATTCAGCG 804

Qy    219 GluArgGluGlu--------------------------------------AspThrLeu 225
Db    805 GAGGAACGGGAATACATCAGCATGTTGAGCACCATGCGGCAACTTGGAGTAACCAGCTTG 864

Qy    226 AsnIleAspMetSerAspIle------TyrAspHisAspProAspLeuTyrAlaLysIle 243
Db    865 AACTTGGATGCGAAGAACCTGAAGGCATACCCATCAACACTCAAGCTATGGCATCAGTTG 924

Qy    244 ValArgTyrProLeuAspIleIleProLeuLeuAspThrGluCysGlnGluValAlaThr 263
Db    925 CATGCTTATCCTCAAGAGATCATTCCATTGATGGATCAGACAGTGAAGGATGTGATGGTG 984

Qy    264 SerLeu------------------------------------------------------ 265
Db    985 GAGCTTGCGATCAAGGAGATGGAGCGCCTGCGGGCTCAGAACCAACGAAACCAGAATCAT 1044

Qy    265 ------------------------------------------------------------ 265
Db   1045 AACAGAGGCCTGAGCTCTGGTCCCGCTGTTCCCAGCTCTGACGCGCTGAGTGAAACTGGC 1104

Qy    266 -----------------LeuProThrPheGluLysHisIleGluAlaArg--------- 276
Db   1105 AGAATGCCACAGAACGAGATCCCCGACCTTGTCGGTGAAGTGGAAACCAAGGCATTCAAA 1164

Qy    277 ------ProPheAsnLeuLysAlaSerValHisMetArgGluLeuAsnProSerAspIle 294
Db   1165 GTTCTGCCTTTTGGTCTCGACTCGACTGTGAATATGAGAGACCTCGATCCTGCAGACATG 1224

Qy    295 AspLysLeuValSerValLysGlyMetValIleArgCysSerSerIleIleProGluIle 314
Db   1225 GATAAACTAGTAAGCATTAAGGGCTTAGTCATTCGAACGACACCCATCATTCCTGATATG 1284

Qy    315 LysGlyAlaPhePheLysCysLeuValCysGlyHisSerProProLeuValThrValVal 334
Db   1285 AAAGAAGCTTTCTTCCGTTGCCAAGTCTGCAACCATGGTGTTCAG---GTCGACATTGAT 1341

Qy    335 LysGlyArgValGluGluProThrArgCysGluLysProGluCysAlaAlaArgAsnAla 354
Db   1342 CGTGGAAAGATTGCGGAACCCACTGAGTGCCCACGTCCAGTGTGTAAGGAACGAAACTCG 1401

Qy    355 MetSerLeuIleHisAsnArgCysThrPheAlaAsnLysGlnIleValArgLeuGlnGlu 374
Db   1402 ATGCAACTCATCCATAACCGCTGTGTATTTGCCGACAAGCAGGTCATCAAGTTGCAGGAA 1461

Qy    375 ThrProAspAlaIleProGluGlyGluThrProHisThrValSerMetCysLeuTyrAsn 394
Db   1462 ACACCTGACAGCATTCCTGATGGCCAGACTCCTCACTCGGTTTCCCTTTGTGTGTATGAT 1521

Qy    395 ThrMetValAspAlaValLysProGlyAspArgIleGluValThrGlyValPheLysAla 414
Db   1522 GAGCTGGTGGATGTCTGCAAGGCTGGTGATCGGGTCGAAGTGACCGGTATTTTCCGGTGC 1581

Qy    415 MetAlaValArgValGlyProAsnGlnArgThrLeuArgAlaLeuTyrLysThrTyrIle 434
Db   1582 AACCCTGTGCGCGTTAATCCTCGCCAGCGTACACAGAAGTCGCTGTTCAAGACGTACATA 1641

Qy    435 AspCysValHisValLysLysSerAspArgGlyArgLeuGlnThrGluAspProMetGlu 454
Db   1642 GATGTTCTTCATGTTCAGAAGATCGATCGCAAGAAGTTGGGTATCGACGTCTCGACCATC 1701

Qy    455 MetAspLysGluAsnAspMetTyrAlaGlyTyrHisGluSerAspThrSerGluAlaAla 474
Db   1702 GAGCAGGAGCTCTCGGAACAGGCGGCTGGGGATGCAGAACAAACACGTAGGCTCACTGCG 1761

Qy    475 AsnGluAlaLysIleGlnLysLeuLysGluLeuSerLysLeuProGlyIleTyrAspArg 494
Db   1762 GAGGAGGAA------GAGAAAATTAAGCGAACTGCTACCAGACCTGATCTGTATGAGCTT 1815

Qy    495 LeuSerArgSerLeuAlaProSerIleTrpGluLeuGluAspIleLysLysGlyLeuLeu 514
Db   1816 CTCTCTCGGTCCTTGGCCCCCAGCATCTACGAGATGGACGACGTGAAGAAGGGAATCCTG 1875

Qy    515 CysGlnLeuPheGlyGlyLysAlaLysLysIleProSerGlyAlaSer------PheArg 532
Db   1876 CTTCAGTTGTTTGGAGGCACCAACAAGACCTTCCAGAAGGGTGGTAACCCACGATACCGT 1935

Qy    533 GlyAspIleAsnValLeuLeuValGlyAspProGlyThrSerLysSerGlnLeuLeuGln 552
```

```
                ||||||||||:::||||||  ||||||||  |||||||||||||||||||:::
Db       1936 GGAGATATCAATATCCTTCTCTGTGGTGACCCATCTACATCCAAGTCCCAGCTTCTTCGT 1995

Qy        553 TyrValHisLysIleAlaProArgGlyIleTyrThrSerGlyArgGlySerSerAlaVal 572
              ||||||||||||||||||||||||||:::||||||||||:::||||||||||||
Db       1996 TACGTCCATAAGATTGCCCCTCGCGGTGTGTATACCAGCGGCAAGGGCTCCTCGGCTGTT 2055

Qy        573 GlyLeuThrAlaTyrValThrLysAspProGluThrArgGluThrValLeuGluSerGly 592
              ||||||||||||||||||||:::||||||||||||||:::   ||||||||||||
Db       2056 GGTCTTACGGCGTACGTCACCCGCGATCCTGAAACCCGCCAGATGGTCCTCGAGTCGGGT 2115

Qy        593 AlaLeuValLeuSerAspArgGlyIleCysCysIleAspGluPheAspLysMetSerAsp 612
              ||||||||||||||||  |||||||||||||||||||||||||||||||:::::
Db       2116 GCCTTGGTTCTTTCAGACGGCGGTATCTGTTGCATCGACGAGTTCGACAAGATGAACGAA 2175

Qy        613 AsnAlaArgSerMetLeuHisGluValMetGluGlnGlnThrValSerValAlaLysGly 632
              :::  ||||||:::||||||||||||||||||||||||||||||||:::|||||
Db       2176 TCCACTCGGTCCGTTCTGCATGAAGTCATGGAACAACAGACAGTATCTATCGCCAAGGCA 2235

Qy        633 GlyIleIleAlaSerLeuAsnAlaArgThrSerValLeuAlaCysAlaAsnProSerGly 652
              ||||||||   :::||||||||||||||||:::||||||  ||||||||   |||
Db       2236 GGCATTATCACTACTTTGAACGCTAGGACCAGCATCCTGGCTTCCGCCAATCCGATCGGT 2295

Qy        653 SerArgTyrAsnAlaArgLeuSerValIleAspAsnIleGlnLeuProProThrLeuLeu 672
              |||||||||||      |||   |||      ||||||   |||||||||||||||
Db       2296 AGCAGGTACAATCCCAACTTGCCCGTTCCTCAAAATATTGACCTTCCGCCTACCTTGCTC 2355

Qy        673 SerArgPheAspLeuIleTyrLeuMetLeuAspLysProAspGluGlnAsnAspArgArg 692
              ||||||||||||:::||||||:::||||||:::  ||||||||   ||||||||||
Db       2356 TCCCGATTCGACTTGGTATACCTCGTGCTGGACCGAGTGGATGAGCAGGAAGATCGTCGG 2415

Qy        693 LeuAlaArgHisLeuValAlaLeuHisTyrGluAsn------TyrGluValSerLysGln 710
              ||||||:::||||||||   ::::::   |||:::            :::::::|||
Db       2416 CTCGCTAAGCATCTTGTCAATATGTACCTGGAAGACAGACCTGAGCATGCTGCCGAGCAA 2475

Qy        711 AspAlaLeuAspLeuGlnThrLeuThrAlaTyrIleThrTyrAlaArgGlnHisValHis 730
              :::   |||   ::::::   ||||||||||||||||||||||:::        ||||||
Db       2476 GAAATCTTGCCGATCGAATTCCTTACAGCCTATATCACCTACGCCAAGACCAAAGTCCAT 2535

Qy        731 ProThrLeuSerAspGluAlaAlaGluAspLeuIleAsnGlyTyrValGluMetArgGln 750
              |||   |||:::         |||   :::   |||   :::   ||||||   ||||||:::
Db       2536 CCAGTGCTCACACCGGCCGCCGGTAAAGCCTTGTCGGATGCTTACGTTAACATGCGTAAG 2595

Qy        751 LysGlyAsnPheProGlySerSerLysLysValIleThrAlaThrProArgGlnLeuGlu 770
                 |||:::         ||||||   :::   ||||||||||||   ||||||||||||
Db       2596 CTTGGAGATGACATCCGGTCTTCTGACCGCCGTATCACCGCTACCACTCGTCAACTGGAG 2655

Qy        771 SerMetIleArgIleSerGluAlaLeuAlaArgMetArgPheSerGluValValGluLys 790
              ||||||||||||:::|||||||||   ||||||||||   |||      |||
Db       2656 TCCATGATCCGACTGTCGGAAGCGCATGCGCGTATGCGGCTATCGCCGGAGGTCACTGCG 2715

Qy        791 ValAspAlaAlaGluAlaValArgLeuLeuAspValAlaLeuGlnGlnSerAlaThrAsp 810
                 |||      |||||||||||||||:::      |||::::::|||:::|||||||||
Db       2716 GATGATGTGGAGGAAGCCGTGCGCCTGATCCGCTCCGCCATCAAGCAGGCGGCCACTGAC 2775

Qy        811 HisAlaThrGlyThrIleAspMetAspLeuIleThrThrGlyValSerAlaSerGluArg 830
                    ||||||   |||||||||   |||:::|||   |||   |||||||||||||||
Db       2776 TCTCGGACCGGTCTGATCGACATGAGCTTGTTGACGGAGGGCACTAGTGCCAGCGAGAGA 2835

Qy        831 IleArgArgAlaAsnLeuLeuAlaAlaLeuArgGluLeuIleAlaAspLysIleSerPro 850
                 |||      |||      ||||||      ::::::   |||      |||
Db       2836 CGCAGCCGGGAAGCACTCAAGCGTGCCTTGCTGTCTGTGGTGGATGATCTGTGCAGCGGT 2895

Qy        851 GlySerSerSerGlyLeuLysThrSerGlnLeuLeuGluAspIleArgSerGlnSerSer 870
              |||   ::::::      :::   :::::::::         :::         ||||||
Db       2896 GGCGGTGCAGCT------CGCTGGGCCGAGGTCTTCAGGATCTTAAGCGAGAACAGCAGC 2949

Qy        871 ValAspValSerLeuGlnAspIleLysAsnAlaLeuGlySerLeuGlnGlyGluGlyPhe 890
              ::::::|||            :::|||:::   :::|||:::   ||||||
Db       2950 ATTGAGGTGGATGGAGCCCAGTTTGCGGATGCGGTTCGAGCGCTGGAGGCTGAGGGAGCG 3009

Qy        891 LeuThrValHisGlyAspIleValLysArg 900
              ::::::|||   |||:::      :::|||
```

```
Db       3010 GTGAGTGTGGTCGGCGAGGGTGCGCGGCGG 3039
```

RESULT 7
US-10-128-714-2558
; Sequence 2558, Application US/10128714
; Publication No. US20030119013A1
; GENERAL INFORMATION:
; APPLICANT: Jiang, Bo
; APPLICANT: Hu, Wenqi
; APPLICANT: Tishkoff, Daniel
; APPLICANT: Zamudio, Carlos
; APPLICANT: Eroshkin, Alexey M
; APPLICANT: Lemieux, Sebastien M
; TITLE OF INVENTION: Identification of Essential Genes in Aspergillus fumigatus and
; TITLE OF INVENTION: Methods of Use
; FILE REFERENCE: 10182-018-999
; CURRENT APPLICATION NUMBER: US/10/128,714
; CURRENT FILING DATE: 2002-04-23
; PRIOR APPLICATION NUMBER: US 60/285,697
; PRIOR FILING DATE: 2001-04-23
; PRIOR APPLICATION NUMBER: US 60/287,066
; PRIOR FILING DATE: 2001-04-27
; PRIOR APPLICATION NUMBER: US 60/295,890
; PRIOR FILING DATE: 2001-06-05
; PRIOR APPLICATION NUMBER: US 60/303,899
; PRIOR FILING DATE: 2001-07-09
; PRIOR APPLICATION NUMBER: US 60/316,362
; PRIOR FILING DATE: 2001-08-31
; NUMBER OF SEQ ID NOS: 8603
; SOFTWARE: PatentIn version 3.1
; SEQ ID NO 2558
;   LENGTH: 2625
;   TYPE: DNA
;   ORGANISM: Aspergillus fumigatus
US-10-128-714-2558

Alignment Scores:

Pred. No.:	1.76e-150	Length:	2625
Score:	1871.50	Matches:	410
Percent Similarity:	63.72%	Conservative:	145
Best Local Similarity:	47.07%	Mismatches:	239
Query Match:	40.61%	Indels:	77
DB:	5	Gaps:	16

US-10-768-511-8 (1-901) x US-10-128-714-2558 (1-2625)

```
Qy         89 ThrProGlyThrProGlyThrProValAlaThrProValTyrAlaThrProValGlyThr 108
              :::||||||   |||         |||   |||         |||::::::
Db         38 AGTCCCGGCGTTCCGACATTCCCAGCAGCAGCAGCGGATTATTCGTCT------------ 85

Qy        109 ProMetGlyThrPro-SerPheHisArgGlyThrPro-GlnTyrLysGlnArgSerGluL 128
                               |||  :::   ||||||      |||          :::|||||||:::|
Db         86 ---------CGTCCAGACCCAGCATCGAGAGCAACCGTGCCGTGTCCCGCCGTAGCGACC 136

Qy        128 euGlySerGlnGly------LysProLeuHisArgArgArgArgSerGlnSerArgGluP 146
               ||   |||   |||                 :::||||||||||||
Db        137 TTCATTCTGGTGGTTTTCTTTCTAGCCCGAACCGTCGCCGCAGAGTTTTTGTCGATGCTA 196

Qy        146 roGlyHisArgSerProSerArgGluProSerAlaAspGlyArgProSerGluSerAla- 165
                 |||                           |||:::   ||||||   |||         |||
Db        197 ATGGT------------------ATGCCTGCGACCGATGGCGATCCACGTTCCGATGCCA 238

Qy        166 --------------GluProAsp------------AspThrLeuGlyGlyGluTyrAla- 176
                            ||||||              :::   |||||||||
Db        239 CCTTCTCGAATATCCACCCAGACACCTCTGAGGCCGAGGCCTTGGGCGGTAGCTCAACCC 298

Qy        177 --TyrValTrpGlyThrAsnValAsnIleProAspValLeuArgAlaIleArgArgPheL 196
                 :::||||||||||||:::::::|||   |||   :::   |||   :::   ||||
Db        299 GTGTGATTTGGGGTACCAACATCTCCATCCAGGATTCCATGTCCGCATTCAAGAACTTTC 358

Qy        196 euHisAsn------------TyrArgSerSerAlaHisAspLeuAsnSerLysTyrIleG 212
               ||:::|||             ||||||      |||                        :
Db        359 TCTACAACTTCCAAACAAAATACCGTCTGTGGGCAGAGGGCGCAACCGAGGATGAGACAC 418
```

```
Qy    212 lnIleIleGluGluThrValGluArgGluGlu--------------------------- 222
Db    419 GTATAATGGGTGATTCAGCGGAGGAACGGGAATACATCAGCATGTTGAGCACCATGCGGC 478

Qy    223 -----------AspThrLeuAsnIleAspMetSerAspIle------TyrAspHisAspP 237
Db    479 AACTTGGAGTAACCAGCTTGAACTTGGATGCGAAGAACCTGAAGGCATACCCATCAACAC 538

Qy    237 roAspLeuTyrAlaLysIleValArgTyrProLeuAspIleIleProLeuLeuAspThrG 257
Db    539 TCAAGCTATGGCATCAGTTGCATGCTTATCCTCAAGAGATCATTCCATTGATGGATCAGA 598

Qy    257 luCysGlnGluValAlaThrSerLeu---------------------------------- 265
Db    599 CAGTGAAGGATGTGATGGTGGAGCTTGCGATCAAGGAGATGGAGCGCCTGCGGGCTCAGA 658

Qy    266 -----------------------------LeuProThrPheGlu---LysHisIleGluA 275
Db    659 ACCAACGAAACCAGAATCATAACAGAGGCCTGAGCTCTGTGGAAACCAAGGCATTCAAAG 718

Qy    275 laArgProPheAsnLeuLysAlaSerValHisMetArgGluLeuAsnProSerAspIleA 295
Db    719 TTCTGCCTTTTGGTCTCGACTCGACTGTGAATATGAGAGACCTCGATCCTGCAGACATGG 778

Qy    295 spLysLeuValSerValLysGlyMetValIleArgCysSerSerIleIleProGluIleL 315
Db    779 ATAAACTAGTAAGCATTAAGGGCTTAGTCATTCGAACGACACCCATCATTCCTGATATGA 838

Qy    315 ysGlyAlaPhePheLysCysLeuValCysGlyHisSerProProLeuValThrValValL 335
Db    839 AAGAAGCTTTCTTCCGTTGCCAAGTCTGCAACCATGGTGTTCAG---GTCGACATTGATC 895

Qy    335 ysGlyArgValGluGluProThrArgCysGluLysProGluCysAlaAlaArgAsnAlaM 355
Db    896 GTGGAAAGATTGCGGAACCCACTGAGTGCCCACGTCCAGTGTGTAAGGAACGAAACTCGA 955

Qy    355 etSerLeuIleHisAsnArgCysThrPheAlaAsnLysGlnIleValArgLeuGlnGluT 375
Db    956 TGCAACTCATCCATAACCGCTGTGTATTTGCCGACAAGCAGGTCATCAAGTTGCAGGAAA 1015

Qy    375 hrProAspAlaIleProGluGlyGluThrProHisThrValSerMetCysLeuTyrAsnT 395
Db   1016 CACCTGACAGCATTCCTGATGGCCAGACTCCTCACTCGGTTTCCCTTTGTGTGTATGATG 1075

Qy    395 hrMetValAspAlaValLysProGlyAspArgIleGluValThrGlyValPheLysAlaM 415
Db   1076 AGCTGGTGGATGTCTGCAAGGCTGGTGATCGGGTCGAAGTGACCGGTATTTTCCGGTGCA 1135

Qy    415 etAlaValArgValGlyProAsnGlnArgThrLeuArgAlaLeuTyrLysThrTyrIleA 435
Db   1136 ACCCTGTGCGCGTTAATCCTCGCCAGCGTACACAGAAGTCGCTGTTCAAGACGTACATAG 1195

Qy    435 spCysValHisValLysLysSerAspArgGlyArgLeuGlnThrGluAspProMetGluM 455
Db   1196 ATGTTCTTCATGTTCAGAAGATCGATCGCAAGAAGTTGGGTATCGACGTCTCGACCATCG 1255

Qy    455 etAspLysGluAsnAspMetTyrAlaGlyTyrHisGluSerAspThrSerGluAlaAlaA 475
Db   1256 AGCAGGAGCTCTCGGAACAGGCGGCTGGGGATGCAGAACAAACACGTAGGCTCACTGCGG 1315

Qy    475 snGluAlaLysIleGlnLysLeuLysGluLeuSerLysLeuProGlyIleTyrAspArgL 495
Db   1316 AGGAGGAA------GAGAAAATTAAGCGAACTGCTACCAGACCTGATCTGTATGAGCTTC 1369

Qy    495 euSerArgSerLeuAlaProSerIleTrpGluLeuGluAspIleLysLysGlyLeuLeuC 515
Db   1370 TCTCTCGGTCCTTGGCCCCCAGCATCTACGAGATGGACGACGTGAAGAAGGGAATCCTGC 1429

Qy    515 ysGlnLeuPheGlyGlyLysAlaLysLysIleProSerGlyAlaSer------PheArgG 533
Db   1430 TTCAGTTGTTTGGAGGCACCAACAAGACCTTCCAGAAGGGTGGTAACCCACGATACCGTG 1489

Qy    533 lyAspIleAsnValLeuLeuValGlyAspProGlyThrSerLysSerGlnLeuLeuGlnT 553
Db   1490 GAGATATCAATATCCTTCTCTGTGGTGACCCATCTACATCCAAGTCCCAGCTTCTTCGTT 1549

Qy    553 yrValHisLysIleAlaProArgGlyIleTyrThrSerGlyArgGlySerSerAlaValG 573
Db   1550 ACGTCCATAAGATTGCCCCTCGCGGTGTGTATACCAGCGGCAAGGGCTCCTCGGCTGTTG 1609

Qy    573 lyLeuThrAlaTyrValThrLysAspProGluThrArgGluThrValLeuGluSerGlyA 593
Db   1610 GTCTTACGGCGTACGTCACCCGCGATCCTGAAACCCGCCAGATGGTCCTCGAGTCGGGTG 1669

Qy    593 laLeuValLeuSerAspArgGlyIleCysCysIleAspGluPheAspLysMetSerAspA 613
Db   1670 CCTTGGTTCTTTCAGACGGCGGTATCTGTTGCATCGACGAGTTCGACAAGATGAACGAAT 1729

Qy    613 snAlaArgSerMetLeuHisGluValMetGluGlnGlnThrValSerValAlaLysGlyG 633
Db   1730 CCACTCGGTCCGTTCTGCATGAAGTCATGGAACAACAGACAGTATCTATCGCCAAGGCAG 1789

Qy    633 lyIleIleAlaSerLeuAsnAlaArgThrSerValLeuAlaCysAlaAsnProSerGlyS 653
Db   1790 GCATTATCACTACTTTGAACGCTAGGACCAGCATCCTGGCTTCCGCCAATCCGATCGGTA 1849

Qy    653 erArgTyrAsnAlaArgLeuSerValIleAspAsnIleGlnLeuProProThrLeuLeuS 673
Db   1850 GCAGGTACAATCCCAACTTGCCCGTTCCTCAAAATATTGACCTTCCGCCTACCTTGCTCT 1909

Qy    673 erArgPheAspLeuIleTyrLeuMetLeuAspLysProAspGluGlnAsnAspArgArgL 693
Db   1910 CCCGATTCGACTTGGTATACCTCGTGCTGGACCGAGTGGATGAGCAGGAAGATCGTCGGC 1969

Qy    693 euAlaArgHisLeuValAlaLeuHisTyrGluAsn------TyrGluValSerLysGlnA 711
Db   1970 TCGCTAAGCATCTTGTCAATATGTACCTGGAAGACAGACCTGAGCATGCTGCCGAGCAAG 2029

Qy    711 spAlaLeuAspLeuGlnThrLeuThrAlaTyrIleThrTyrAlaArgGlnHisValHisP 731
Db   2030 AAATCTTGCCGATCGAATTCCTTACAGCCTATATCACCTACGCCAAGACCAAAGTCCATC 2089

Qy    731 roThrLeuSerAspGluAlaAlaGluAspLeuIleAsnGlyTyrValGluMetArgGlnL 751
Db   2090 CAGTGCTCACACCGGCCGCCGGTAAAGCCTTGTCGGATGCTTACGTTAACATGCGTAAGC 2149

Qy    751 ysGlyAsnPheProGlySerSerLysLysValIleThrAlaThrProArgGlnLeuGluS 771
Db   2150 TTGGAGATGACATCCGGTCTTCTGACCGCCGTATCACCGCTACCACTCGTCAACTGGAGT 2209

Qy    771 erMetIleArgIleSerGluAlaLeuAlaArgMetArgPheSerGluValValGluLysV 791
Db   2210 CCATGATCCGACTGTCGGAAGCGCATGCGCGTATGCGGCTATCGCCGGAGGTCACTGCGG 2269

Qy    791 alAspAlaAlaGluAlaValArgLeuLeuAspValAlaLeuGlnGlnSerAlaThrAspH 811
Db   2270 ATGATGTGGAGGAAGCCGTGCGCCTGATCCGCTCCGCCATCAAGCAGGCGGCCACTGACT 2329

Qy    811 isAlaThrGlyThrIleAspMetAspLeuIleThrThrGlyValSerAlaSerGluArgI 831
Db   2330 CTCGGACCGGTCTGATCGACATGAGCTTGTTGACGGAGGGCACTAGTGCCAGCGAGAGAC 2389

Qy    831 leArgArgAlaAsnLeuLeuAlaAlaLeuArgGluLeuIleAlaAspLysIleSerProG 851
Db   2390 GCAGCCGGGAAGCACTCAAGCGTGCCTTGCTGTCTGTGGTGGATGATCTGTGCAGCGGTG 2449

Qy    851 lySerSerSerGlyLeuLysThrSerGlnLeuLeuGluAspIleArgSerGlnSerSerV 871
Db   2450 GCGGTGCAGCT------CGCTGGGCCGAGGTCTTCAGGATCTTAAGCGAGAACAGCAGCA 2503

Qy    871 alAspValSerLeuGlnAspIleLysAsnAlaLeuGlySerLeuGlnGlyGluGlyPheL 891
Db   2504 TTGAGGTGGATGGAGCCCAGTTTGCGGATGCGGTTCGAGCGCTGGAGGCTGAGGGAGCGG 2563

Qy    891 euThrValHisGlyAspIleValLysArg 900
```

```
               :::::|||   |||:::      :::|||
Db       2564 TGAGTGTGGTCGGCGAGGGTGCGCGGCGG 2592

RESULT 8
US-10-425-115-98897
; Sequence 98897, Application US/10425115
; Publication No. US20040214272A1
; GENERAL INFORMATION:
;  APPLICANT: La Rosa, Thomas J.
;  APPLICANT:  Kovalic, David K.
;  APPLICANT:  Zhou, Yihua
;  APPLICANT:  Cao, Yongwei
;  TITLE OF INVENTION: Nucleic Acid Molecules and Other Molecules Associated With
;  TITLE OF INVENTION:  Plants
;  FILE REFERENCE: 38-21(53222)B
;  CURRENT APPLICATION NUMBER: US/10/425,115
;  CURRENT FILING DATE:  2003-04-28
;  NUMBER OF SEQ ID NOS: 369326
; SEQ ID NO 98897
;   LENGTH: 2939
;   TYPE: DNA
;   ORGANISM: Zea mays
;   FEATURE:
;   OTHER INFORMATION: Clone ID: MRT4577_21707C.1
US-10-425-115-98897

Alignment Scores:
Pred. No.:              6.57e-149      Length:         2939
Score:                  1854.00        Matches:        426
Percent Similarity:     62.34%         Conservative:   145
Best Local Similarity:  46.51%         Mismatches:     244
Query Match:            40.23%         Indels:         102
DB:                     8              Gaps:           24

US-10-768-511-8 (1-901) x US-10-425-115-98897 (1-2939)

Qy         20 SerGluGlyValSerThrProLeuProGlnValThrSerProSerPheAspAsnAlaAla 39
              |||:::|||       :::       |||       :::||||||    :::        :::
Db        113 TCTGACGGCGGCGGCAGCTCCCCTCCTCCCGCATCGTCGCCC---TATGGTCGGCCGTCA 169

Qy         40 SerProValAlaGlyArgArgAlaValArgGlnThrProThrSerAlaValArgArgArg 59
              ||||||:::|||                              :::|||:::      ||||||
Db        170 AGCCCACTCGCG---------GTTACTAACTCCTCCCCATCTCAGCCCACACGCCGCTCC 220

Qy         60 GlyArgGluThrAspSerAlaArgArgArgArgSerArgSerArgSerLeuGlyAsnSer 79
              |||               ||||||||||||
Db        221 GGC---------------GGTCGCCGTCGCCGTGGTTCC--------------------- 244

Qy         80 ValTyrSerSerProTyrAspAlaGlyThrProGlyThrProGlyThrProValAlaThr 99
                    :::||||||||   |||   :::|||         |||
Db        245 ------GCTAGTCCTTAT---GCTTCGTCCCCTTCTCTCGGGGGG--------------- 280

Qy        100 ProValTyrAlaThrProValGlyThrProMetGlyThrProSerPheHisArgGlyThr 119
                    :::   ||||||                        |||      |||   |||
Db        281 ------TTCGAGACGCCG--------------------CCGCACCCTGGCCGCCGCACG 313

Qy        120 ProGlnTyrLysGlnArgSerGluLeuGlySerGlnGlyLysProLeuHisArgArgArg 139
              |||                     |||      |||:::   |||   |||         |||:::
Db        314 CCG---------------TCCGGAGCTGGGGCT---GGCGCCCCTCGGCAGCCGCGTCAG 355

Qy        140 ArgSerGlnSerArgGluProGlyHisArgSerProSerArgGluProSerAlaAspGly 159
                 |||      |||   |||                 :::||||||      |||   |||
Db        356 AACTCGACTGGCCGGTTCCCG------CCGACGCCCTCCACTCCAATGTCCACCGATGAC 409

Qy        160 ---ArgProSerGluSerAlaGluProAspAspThrLeuGlyGly--------------- 173
                    ||||||      |||      |||:::|||   ||||||
Db        410 GTCCCTCCATCCTCCGAAGCTGGGGACGACGAGACCGACGGCGGCGGCGGCGTCGACGCC 469

Qy        174 GluTyrAlaTyrValTrpGlyThrAsnValAsnIleProAspValLeuArgAlaIleArg 193
                       :::||||||||||||||||:::::::::   ||||||      ||||||
Db        470 ACCCCGGTCTTCGTCTGGGGCACCAACATAAGCGTGCAGGACGTCAACGCCGCCATTCTC 529

Qy        194 ArgPheLeuHisAsnTyrArgSerSerAlaHisAspLeu--------------------- 206
              |||||||||   ::::::|||                     |||
Db        530 CGGTTCTTGCGCCACTTCCGG---GACCCGCGCGACGCTGGCCGCGTCGACCCGGTCATG 586

Qy        207 ---AsnSerLysTyrIleGlnIleIleGluGluThrValGluArgGlu---GluAspThr 224
                         ||||||::::::   |||         :::|||   |||      ::::::
Db        587 GACGAGGGCAAGTACATGCGCGCCATCCACCGCATCCTCGAGCTCGAGGGCGGGGAGTCG 646

Qy        225 LeuAsnIleAspMetSerAspIleTyrAspHisAspProAspLeuTyrAlaLysIleVal 244
              |||::::::|||         |||::::::||||||||||||||||||||||:::|||:::|||
Db        647 CTCGACGTCGATGCGCACGACGTGTTTGACCACGACCCAGACCTCTACAGCAAGATGGTT 706

Qy        245 ArgTyrProLeuAspIleIleProLeuLeuAspThrGluCysGlnGluValAlaThrSer 264
              |||||||||||||||:::::::::   :::   |||            ::::::
Db        707 CGCTATCCGCTCGAGGTGCTCGCCATCTTCGACATCGTGCTCATGGACCTCGTCGCGCGC 766

Qy        265 LeuLeuProThrPheGluLysHisIleGluAlaArgProPheAsnLeuLysAlaSerVal 284
              :::   |||   ||||||||||||||||:::      |||   :::|||||||||:::|||:::
Db        767 ATCGAGCCGCTCTTCGAGAAGCACATCCAGACCAGGATCTACAACCTCAAGTCGTCCATT 826

Qy        285 HisMetArgGluLeuAsnProSerAspIleAspLysLeuValSerValLysGlyMetVal 304
                 :::|||   ||||||||||||||||||:::|||:::||||||:::|||||||||:::
Db        827 TGCTTGAGGAATCTCAACCCATCTGATATTGAGAAGATGGTATCCATCAAGGGTATGATA 886

Qy        305 IleArgCysSerSerIleIleProGluIleLysGlyAlaPhePheLysCysLeuValCys 324
              ||||||||||||||||:::|||||||||:::|||   |||   |||:::||||||||||||
Db        887 ATTAGATGCAGCTCGGTCATACCGGAGCTCAAGGAGGCTGTGTTCCGCTGCCTGGTTTGT 946

Qy        325 GlyHisSerProProLeuValThrValValLysGlyArgValGluGluProThrArgCys 344
              |||                  |||   |||   :::||||||||||   ||||||      |||
Db        947 GGTTTCTACTCAGAGCCCGTCATGGTTGATAGAGGAAGAGTAACTGAACCACACATTTGT 1006

Qy        345 GluLysProGluCysAlaAlaArgAsnAlaMetSerLeuIleHisAsnArgCysThrPhe 364
              :::|||   :::|||   |||   |||:::|||:::|||:::||||||||||||   |||
Db       1007 CAGAAAGAACAATGTAAAGCCACAAATTCTATGACCCTAGTGCACAACAGATGCAGATTT 1066

Qy        365 AlaAsnLysGlnIleValArgLeuGlnGluThrProAspAlaIleProGluGlyGluThr 384
              ::::::||||||||||:::::::||||||||||||||||||   ||||||||||||   |||
Db       1067 TCAGACAAGCAGATCATAAAGTTGCAGGAAACACCAGACGAGATACCAGAAGGTGGCACT 1126

Qy        385 ProHisThrValSerMetCysLeuTyrAsnThrMetValAspAlaValLysProGlyAsp 404
              |||||||||||||||:::   :::::::::   :::|||||||||   ||||||||||||
Db       1127 CCACATACAGTTAGTGTCTTGATGCATGATAAGCTTGTTGATGCTGGAAAGCCTGGAGAT 1186

Qy        405 ArgIleGluValThrGlyValPheLysAlaMetAlaValArgValGlyProAsnGlnArg 424
              |||:::|||:::||||||:::::::::||||||::::::|||:::||||||   ||||||
Db       1187 AGGGTTGAGATAACTGGAATATACAGAGCTATGAGTATTCGAATTGGACCAACTCAAAGG 1246

Qy        425 ThrLeuArgAlaLeuTyrLysThrTyrIle-----------------AspCysValHis 438
              |||:::::::::::::::|||||||||||                    :::   |||
Db       1247 ACAGTGAAGTCTATATTCAAGACATATATTGCCCGGATGCGGTTCTCTGAAGTGGTCGAG 1306

Qy        439 ValLysLysSerAspArgGlyArgLeuGlnThrGluAspProMetGluMetAspLysGlu 458
              |||:::                             |||      |||:::
Db       1307 GTGCGAGATGTTGTGGAGGCCTTCAGGCTTCTTGAAGTTGCCATGCAG---CAGTCTGCA 1363

Qy        459 AsnAspMetTyrAlaGlyTyrHisGluSerAsp---------ThrSerGluAlaAlaAsn 475
                 |||      |||      :::   |||         |||      :::|||:::
Db       1364 ACGGATCATGCAACTGGGACGATTGATATGGATCTAATCATGACGGGGATATCTGCAAGT 1423

Qy        476 Glu---------------------------AlaLysIleGlnLysLeuLysGluLeuSer 486
              |||                           |||            :::   |||
Db       1424 GAAAGGCAGAGACGGAACGACCTTGTTGCTGCAACCCGTAACCTGATCGTGGAGAAAATG 1483

Qy        487 LysLeuProGlyIleTyrAspArgLeuSerArgSerLeuAlaProSerIleTrpGluLeu 506
              :::|||   |||                              ||||||:::      :::
Db       1484 CAGCTCGGAGGG-----------------------------CCTTCAATGCATATGATT 1513

Qy        507 GluAspIleLysLysGlyLeuLeuCysGlnLeuPheGlyGlyLysAlaLysLysIlePro 526
              |||   :::   :::         :::   :::|||   |||         :::   :::||||||
Db       1514 GAGTTGCTGTA-AGAGCTCAGGTCGCGGAGCTGAATGGAGAGCCATCTGCACGAATTCCT 1572

Qy        527 SerGlyAlaSerPheArgGlyAspIleAsnValLeuLeuValGlyAspProGlyThrSer 546
                 ||||||                  |||                  ||||||
```

```
Db    1573 CGGGGCGCTTGGCACTATGATGACAGAAGG------TGTGCTTCTGATCCATGGAAAAAA 1626

Qy     547 LysSerGlnLeuLeuGlnTyrValHisLysIleAlaProArgGlyIleTyrThrSerGly 566

Db    1627 GGTGAAGAGAGGTTGATGTTAACT-----------------GCCATCTGCAGGGATGGT 1668

Qy     567 ArgGlySerSerAlaValGlyLeuThrAlaTyr-ValThrLysAspProGluThrArgGl 586

Db    1669 TGTGTGAGCTCC---------------GTCTACCGTTTCCTTCAACAGTCAAGGTTGCGA 1713

Qy     586 uThrValLeuGluSerGlyAlaLeuValLeuSerAspArgGlyIleCysCysIleAspGl 606

Db    1714 AGATCTGATCGACTGTAGGGCATTTGCTTTCAGAGTGCTAAGTAGTTTGTGCTGAGATGA 1773

Qy     606 uPheAspLysMetSerAspAsnAlaArgSerMetLeuHisGluValMetGluGlnGlnTh 626

Db    1774 GTTTCATAAGATGTCAGTTAATGCCCGAAGCATGTTACACGAGGTGATGGAACAGCAGAC 1833

Qy     626 rValSerValAlaLysGlyGlyIleIleAlaSerLeuAsnAlaArgThrSerValLeuAl 646

Db    1834 AGTATCCATTGCGAAGGCTGGAATAATTGCATCTTTAAACGCTAGGACATCTGTCCTGGC 1893

Qy     646 aCysAlaAsnProSerGlySerArgTyrAsnAlaArgLeuSerValIleAspAsnIleGl 666

Db    1894 ATGTGCCAATCCTACTGAATCACGTTACAATCCAAGGCTCTCTGTAATTGACAACATCCA 1953

Qy     666 nLeuProProThrLeuLeuSerArgPheAspLeuIleTyrLeuMetLeuAspLysProAs 686

Db    1954 CTTAGCGCCAACGCTACTTTCAAGATTCGACCTGATTTATCTTATCTGGACAAGGCGGA 2013

Qy     686 pGluGlnAsnAspArgArgLeuAlaArgHisLeuValAlaLeuHisTyrGluAsnTyrGl 706

Db    2014 TGAGCAAACTGATAGGCGCCTGGCAAAGCATATTGTTTCGTTGCATTTTGAGAATCCAAA 2073

Qy     706 uValSerLysGlnAspAlaLeuAspLeuGlnThrLeuThrAlaTyrIleThrTyrAlaAr 726

Db    2074 TTTAGAGGAGCTCGAGGTCTTGGACTTGCAGACACTAGTTTCCTACATAAGCTATGCAAG 2133

Qy     726 gGlnHisValHisProThrLeuSerAspGluAlaAlaGluAspLeuIleAsnGlyTyrVa 746

Db    2134 GAAGTATATTCAGCCACAGTTATCTGATGAAGCTGCAGAAGAGTTAACTCGTGGCTATGT 2193

Qy     746 lGluMetArgGlnLysGlyAsnPheProGlySerSerLysLysValIleThrAlaThrPr 766

Db    2194 CGAGATGAGAAAAAGAGGGAATAGCCCTGGGAGCAGAAAGAAGGTCATAACAGCAACCGC 2253

Qy     766 oArgGlnLeuGluSerMetIleArgIleSerGluAlaLeuAlaArgMetArgPheSerGl 786

Db    2254 TAGACAAATAGAGAGTTTGATCCGTCTCAGCGAAGCATTAGCCCGAATGCGGTTCTCTGA 2313

Qy     786 uValValGluLysValAspAlaAlaGluAlaValArgLeuLeuAspValAlaLeuGlnGl 806

Db    2314 AGTGGTCGAGGTGCGGGATGTTGTGGAGGCATTCAGGCTTCTTGAAGTCGCCATGCAGCA 2373

Qy     806 nSerAlaThrAspHisAlaThrGlyThrIleAspMetAspLeuIleThrThrGlyValSe 826

Db    2374 GTCTGCGACGGATCATGCAACTGGTACGATTGATATGGATCTGATCATGACGGGGATATC 2433

Qy     826 rAlaSerGluArgIleArgArgAlaAsnLeuLeuAlaAlaLeuArgGluLeuIleAlaAs 846

Db    2434 CGCAAGCGAAAGGCAGAGGCGGGAGAACCTCGTTGCCGCAACCCGTAACCTGATTGCGGA 2493

Qy     846 pLysIleSerProGlySerSerSerGlyLeuLysThrSerGlnLeuLeuGluAspIleAr 866

Db    2494 GAAAATGCAGCTTGGAGGCCCCTCG---ATGCGCATGATTGAGTTGCTGGAGGAACTGAG 2550

Qy     866 gSerGlnSerSerValAspValSerLeuGlnAspIleLysAsnAlaLeuGlySerLeuGl 886

Db    2551 GAAGCAGAGCTCAATGGAAATTCATATGCACGAACTCCGCGGTGCTCTTGGCACCCTGAT 2610

Qy     886 nGlyGluGlyPheLeuThrValHisGlyAspIleValLysArgVal 901

Db    2611 GACTGAAGGCGCGGTGGTTATCCATGGAGACAACGTGAGGAGAGTT 2656
```

```
RESULT 9
US-10-260-238-580
; Sequence 580, Application US/10260238
; Publication No. US20040016025A1
; GENERAL INFORMATION:
;  APPLICANT: Budworth, Paul R.
;  APPLICANT:  Moughamer, Todd G.
;  APPLICANT:  Briggs, Steven P.
;  APPLICANT:  Cooper, Bret
;  APPLICANT:  Glazebrook, Jane
;  APPLICANT:  Goff, Stephen A.
;  APPLICANT:  Katagiri, Fumiyaki
;  APPLICANT:  Kreps, Joel
;  APPLICANT:  Provart, Nicholas
;  APPLICANT:  Ricke, Darrell
;  APPLICANT:  Zhu, Tong
;  TITLE OF INVENTION: PROMOTERS FOR REGULATION OF PLANT EXPRESSION
;  FILE REFERENCE: 60111-NP
;  CURRENT APPLICATION NUMBER: US/10/260,238
;  CURRENT FILING DATE:  2002-09-26
;  PRIOR APPLICATION NUMBER: US 60/325,448
;  PRIOR FILING DATE: 2001-09-26
;  PRIOR APPLICATION NUMBER: US 60/325,277
;  PRIOR FILING DATE: 2001-09-26
;  PRIOR APPLICATION NUMBER: US 60/370,620
;  PRIOR FILING DATE: 2002-04-04
;  NUMBER OF SEQ ID NOS: 6077
; SEQ ID NO 580
;   LENGTH: 1723
;   TYPE: DNA
;   ORGANISM: Oryza sativa
;   FEATURE:
;   NAME/KEY: N_region
;   LOCATION: (95)..(95)
;   OTHER INFORMATION: n = any nucleotide
US-10-260-238-580

Alignment Scores:
Pred. No.:              5.1e-148      Length:        1723
Score:                  1840.00       Matches:       403
Percent Similarity:     64.41%        Conservative:  82
Best Local Similarity:  53.52%        Mismatches:    74
Query Match:            39.92%        Indels:        196
DB:                     7             Gaps:          12

US-10-768-511-8 (1-901) x US-10-260-238-580 (1-1723)

Qy     135 LeuHisArgArgArgArgSerGlnSerArgGluProGlyHisArgSerProSerArgGlu 154

Db       3 GTCCACCGACGACGCCC-GGTTCTCCTCGGAGGCCGGGGACGAGGACACCC--------- 52

Qy     155 ProSerAlaAspGlyArgProSerGluSerAlaGluProAspAspThrLeuGlyGlyGlu 174

Db      53 CCGAG-ACCGACGGCGGCGGCGGCGGCGGCGCGGGCGCCGACGNCACG------------ 99

Qy     175 TyrAlaTyrValTrpGlyThrAsnValAsnIleProAspValLeuArgAlaIleArgArg 194

Db     100 CCGGTGTTCGTCTGGGGCACCAACATCAGCGTGCAGGACGTCAACGCCGCCATCCTCCGG 159

Qy     195 PheLeuHisAsnTyrArgSerSerAlaHisAspLeuAsnSerLysTyrIleGlnIleIle 214

Db     160 TTCCTCCGCCACTTCCGGACCCGCGCG------------------------------- 186

Qy     215 GluGluThrValGluArgGluGluAspThrLeuAsnIleAspMetSerAspIleTyrAsp 234

Db     186 ------------------------------------------------------------ 186

Qy     235 HisAspProAspLeuTyrAlaLysIleValArgTyrProLeuAspIleIleProLeuLeu 254

Db     187 ---------------------------------------------ACGCCGGCCGCG 198

Qy     255 AspThrGluCysGlnGluValAlaThrSerLeuLeuProThrPheGluLysHisIleGlu 274
```

```
Db     199 TCGACC---------------------------------------------------------- 204

Qy     275 AlaArgProPheAsnLeuLysAlaSerValHisMetArgGluLeuAsnProSer-AspIl 294

Db     205 ------------------------------------CGGTCATGGACGAGGGCAAGATAT 228

Qy     294 eAspLysLeuValSerValLysGlyMetValIleArgCysSerSerIleIleProGluIl 314

Db     229 CGAGAAGATGGTGTCCATCAAGGGTATGATAATTCGGTGCAGCTCGGTGATTCCAGAGCT 288

Qy     314 eLysGlyAlaPhePheLysCysLeuValCysGly---HisSerProProLeuValThr-V 333

Db     289 CAAGGAGGCTGTGTTCCGCTGCTTGGTTTGCGGGTTCTACTCTGAGCCTGTAATGGTTGA 348

Qy     333 alValLysGlyArgValGluGluProThrArgCysGluLysProGluCysAlaAlaArgA 353

Db     349 TC---------------------------------------------------------GAGA 354

Qy     353 snAlaMetSerLeuIleHisAsnArgCys--ThrPheAlaAsnLysGlnIleValArgLe 372

Db     355 ACTCTATGACTCTTGTGCATAACCGATGCAGAGGTTTGCAGATAAGCAGATCATAAAGTT 414

Qy     372 uGlnGluThrProAspAlaIleProGluGlyGluThrProHisThrValSerMetCysLe 392

Db     415 GCAGGAAACACCAGATGAGATACCAGAAGGTGGCACTCCACATACTGTTAGTGTCTTGAT 474

Qy     392 uTyrAsnThrMetValAspAlaValLysProGlyAspArgIleGluValThrGlyValPh 412

Db     475 GCACGATAAGCTTGTAGATGCTGGAAAGCCTGGAGACAGGGTTGAG---ACTGGGATATA 531

Qy     412 eLysAlaMetAlaValArgValGlyProAsnGlnArgThrLeuArgAlaLeuTyrLysTh 432

Db     532 CAGAGCCATGAGTATTAGAGTTGGCCCAACTCAGAGGACAGTGAAGTCGATTTTCAAGAC 591

Qy     432 rTyrIleAspCysValHisValLysLysSerAspArgGlyArgLeuGlnThrGluAspPr 452

Db     592 GTACATTGATTGCCTTCACATAAAG----------------------------------- 616

Qy     452 oMetGluMetAspLysGluAsnAspMetTyrAlaGlyTyrHisGluSerAspThrSerGl 472

Db     616 ------------------------------------------------------------ 616

Qy     472 uAlaAlaAsnGluAlaLysIleGlnLysLeuLysGluLeuSerLysLeuProGlyIleTy 492

Db     617 ------------------------------AAGGAGTTGTCAAAGTTGCCAGATATATA 645

Qy     492 rAspArgLeuSerArgSerLeuAlaProSerIleTrpGluLeuGluAspIleLysLysGl 512

Db     646 TGACAGATTAACTAGGTCATTGGCTCCAAACATATGGGAGCTGGACGACGTTAAAAGGGG 705

Qy     512 yLeuLeuCys---GlnLeuPheGlyGlyLysAlaLysLysIleProSerGlyAlaSerPh 531

Db     706 CCTCCTTTGCCAGCAGCTTTTTGGTGGAAATGCTTTGAGGCTTCCTTCTGGAGCTAGTTT 765

Qy     531 eArgGlyAspIleAsnValLeuLeuValGlyAspProGlyThrSerLysSerGlnLeuLe 551

Db     766 CCGAGGCGACATCAATATTTTGCTTGTTGGTGATCCTGGAACAAGTAAATCCCAGCTTCT 825

Qy     551 uGlnTyrValHisLysIleAlaProArgGlyIleTyrThrSerGlyArgGlySerSerAl 571

Db     826 CCAATACATGCACAAACTGTCTCCTCGTGGCATTTATACAAGTGGCAGAGGAAGTTCAGC 885

Qy     571 aValGlyLeuThrAlaTyrValThrLysAspProGluThrArgGluThrValLeuGluSe 591

Db     886 TGTTGGCCTTACTGCATACGTTACCAAGGATCCTGAAACTGGTGAAACTGTTCTTGAGAG 945

Qy     591 rGlyAlaLeuValLeuSerAspArgGlyIleCysCysIleAspGluPheAspLysMetSe 611

Db     946 TGGAGCGCTTGTTTTGAGTGACAAAGGTGTTTGTTGTATTGATGAATTTGATAAGATGTC 1005

Qy     611 rAspAsnAlaArgSerMetLeuHisGluValMetGluGlnGlnThrValSerValAlaLy 631

Db    1006 TGATAATGCTCGAAGCATGTTACATGAGGTGATGGAACAACAGACTGTCTCCATTGCCAA 1065

Qy     631 sGlyGlyIleIleAlaSerLeuAsnAlaArgThrSerValLeuAlaCysAlaAsnProSe 651

Db    1066 GGCTGGAATAATTGCATCTTTAAATGCCAGAACATCAGTTCTAGCATGTGCAAATCCTAC 1125

Qy     651 rGlySerArgTyrAsnAlaArgLeuSerValIleAspAsnIleGlnLeuProProThrLe 671

Db    1126 TGAATCACGTTATAATCCAAGGCTCTCTGTGATTGACAATATCCATCTTCCTCCAACACT 1185

Qy     671 uLeuSer--ArgPheAspLeuIleTyrLeuMetLeuAspLysProAspGluGlnAsnAsp 690

Db    1186 GCTTTCTAGAGGTTTGACCTCATTTATCTGATATTGGACAAGGCAGATGAGCAAACTGAT 1245

Qy     691 ArgArgLeuAlaArgHisLeuValAlaLeuHisTyrGluAsnTyrGluValSerLysGln 710

Db    1246 AGACGCCTGGCTAAGCATATTGTTTCGTTGCATTTTGAGAAT------------------ 1287

Qy     711 AspAlaLeuAspLeuGlnThrLeuThrAlaTyrIleThrTyrAlaArgGlnHisValHis 730

Db    1288 -----------------------CCAAACTACATAAGTTATGCAAGGAAGCATATACAA 1323

Qy     731 ProThrLeuSerAspGluAlaAlaGluAspLeuIleAsnGlyTyrValGluMetArgGln 750

Db    1324 CCACAGTTATCTGATGAAGCTGCAGAAGAATTGACTCGCGGCTATGTTGAGATGAGGAAA 1383

Qy     751 LysGlyAsnPheProGlySerSerLysLysValIleThrAlaThrProArgGlnLeuGlu 770

Db    1384 AGAGGAAACAGCCCTGGTAGCAGAAAGAAG---ATAACTGCGACAGCTCGACAAATTGAG 1440

Qy     771 SerMetIleArgIleSerGluAlaLeuAlaArgMetArgPheSerGluValValGluLys 790

Db    1441 AGCTTGATTCGGCTCAGTGAAGCACTGGCCCGAATGCGATTCTCTGAAATGGTTGAAGTA 1500

Qy     791 ValAspAlaAlaGluAlaValArgLeuLeuAspValAlaLeuGlnGlnSerAlaThrAsp 810

Db    1501 CAAGATGTTGTAGAGGCCTTCAGGCTTCTCGAAGTTGCCATGCAGCAATCGGCAACTGAT 1560

Qy     811 HisAlaThr-GlyThrIleAspMetAspLeuIleThrThrGlyValSerAlaSerGluAr 830

Db    1561 CATGCCACTGGGTACAATCGATATGGATCTTATCATGACTGGAATATCTGCGAGCGAAAG 1620

Qy     830 gIleArgArgAlaAsnLeuLeuAlaAlaLeuArgGluLeuIleAlaAspLysIleSerPr 850

Db    1621 GCAGAGGCGG-------------------------------------------------- 1630

Qy     850 oGlySerSerSerGlyLeuLysThrSerGlnLeuLeuGluAspIleArgSerGlnSerSe 870

Db    1631 ------------------------GACCAGTTGCTGGAAGAAATTAGGAAGCAGAGCTC 1665

Qy     870 rValAspValSerLeuGlnAspIle 878

Db    1666 TATGGAAGTTCATCTGCATGATGTA 1690
```

RESULT 10
US-10-320-797-2332
; Sequence 2332, Application US/10320797
; Publication No. US20040014955A1
; GENERAL INFORMATION:
; APPLICANT: Eroshkin, Alexey M.
; APPLICANT: Zamudio, Carlos
; TITLE OF INVENTION: IDENTIFICATION OF ESSENTIAL GENES OF CRYPTOCOCCUS NEOFORMANS AND
; TITLE OF INVENTION: METHODS OF USE
; FILE REFERENCE: 10182-021-999
; CURRENT APPLICATION NUMBER: US/10/320,797
; CURRENT FILING DATE: 2002-12-16
; PRIOR APPLICATION NUMBER: 60/341,261
; PRIOR FILING DATE: 2001-12-17
; NUMBER OF SEQ ID NOS: 3361
; SOFTWARE: PatentIn version 3.1
; SEQ ID NO 2332
;   LENGTH: 2970
;   TYPE: DNA

```
;   ORGANISM: Cryptococcus neoformans
US-10-320-797-2332

Alignment Scores:
Pred. No.:              2.6e-147      Length:        2970
Score:                  1835.50       Matches:       409
Percent Similarity:     59.59%        Conservative:  147
Best Local Similarity:  43.84%        Mismatches:    246
Query Match:            39.82%        Indels:        131
DB:                     7             Gaps:          20

US-10-768-511-8 (1-901) x US-10-320-797-2332 (1-2970)

Qy     47 AlaValArgGlnThrProThrSerAlaValArgArg--------------ArgGlyArg 61
Db     10 GCTCTGCTGCAGTCAGACGAATCCGCGATGAAACACCATTGTTCTTCCCTCCGTGGTAGT 69

Qy     62 GluThrAspSerAlaArgArgArgSerArgSerArgSerLeuGlyAsnSerValTyr 81
Db     70 TCCGGAGGATCGACCCCTCGTCGAGCTCGTCGAGGTGACATCCACTCCTCTTTCCCGCAC 129

Qy     82 SerSerPro---------------------------------------------TyrAsp 86
Db    130 TCATCGCCGTCTCTCGCTCGACGACAAGCCCAAGTCCCGGCTGACTCTCTTTCGCTCCGT 189

Qy     87 AlaGlyThrProGlyThrProGlyThrProValAlaThrProValTyrAlaThrProVal 106
Db    190 GCGTCTTCCCCTGGT--------------------------------------------- 204

Qy    107 GlyThrProMetGlyThrProSerPheHisArgGlyThrProGlnTyrLysGlnArgSer 126
Db    205 ---------ATGGGTACCGATAGCTTT------GGTACTCCGCGTGTT------------ 237

Qy    127 GluLeuGlySerGlnGlyLysProLeuHisArgArgArgArgSerGlnSerArgGluPro 146
Db    238 ---TTCGGATCGGCTGCCCCTACCCTTTCCGCAGTTGCCCAATCTCAACTT--------- 285

Qy    147 GlyHisArgSerProSerArgGluProSerAlaAspGlyArgProSerGluSerAlaGlu 166
Db    286 --------------------------GGTGCCGACGGC---------------------- 297

Qy    167 ProAspAspThrLeuGlyGlyGluTyrAlaTyrValTrpGlyThrAsnValAsnIlePro 186
Db    298 ------GACGATATCGATGGTATGGTCAAATTTATCTGGGGTACGACAATCTCTTTGCAA 351

Qy    187 AspValLeuArgAlaIleArgArgPheLeu------------------------------ 196
Db    352 GAGAGCATGAACCTTTTCCGGGATTTTTTGAGAGGCTTCAAACCCAAGTACCGTGCTGTG 411

Qy    197 HisAsnTyrArgSerSerAlaHisAspLeuAsnSer------------------------ 208
Db    412 TACAATGCGGAACAATCCAGACATGCTGAGGAGAGTGGCGGCGTTGCGCCTCCTCCGATG 471

Qy    209 ---LysTyrIleGlnIleIleGluGluThrValGlu------------------------ 219
Db    472 ACTTTGTACGACAATCTCTCTGCTGAGCGTGCCGAGGTCCCTCTCTATGAGACGTATCTC 531

Qy    220 ---------------ArgGluGluAspThrLeuAsnIleAspMetSerAspIleTyrAsp 234
Db    532 AACCGGCTTCGACTCACTGGTGAAACTAATTTAAACCTCGACGCGCTCAATCTGCTCGCG 591

Qy    235 HisAspPro------AspLeuTyrAlaLysIleValArgTyrProLeuAspIleIlePro 252
Db    592 TACCGACCCACAAAAAAGCTTTACCAGCAATTGGTCAACTATCCGCAAGAAGTGATCCCT 651

Qy    253 LeuLeuAspThrGluCysGlnGluValAlaThr--------------------------- 263
Db    652 ATCATGGATCAGGTCTTGAGGGATGTGATGATTGAATTGGGTCATGAAGAGCTGGAGAAA 711

Qy    264 ------------------------SerLeuLeuProThrPheGluLysHisIle---Glu 274
Db    712 GCGCAGACCAAATTCGCTGAGGGTAACTTATCTCAAGATGTGGAGTCACGAGTGTATAAA 771

Qy    275 AlaArgProPheAsnLeuLysAlaSerValHisMetArgGluLeuAsnProSerAspIle 294
Db    772 GTCAGACCGTTTGGTGGAGAGAAGACGGTCAACATGAGGGATCTTAATCCCGGAGATACT 831

Qy    295 AspLysLeuValSerValLysGlyMetValIleArgCysSerSerIleIleProGluIle 314
Db    832 GACAAGCTGGTGACCGTGAAGGGTCTTGTCATCCGTGCAACTCCCGTCATTCCTGATATG 891

Qy    315 LysGlyAlaPhePheLysCysLeuValCysGlyHisSerProProLeuValThrValVal 334
Db    892 ACCACTGCTTCTTCCGATGCTTGGTTTGTCAGCACACTGTTCAA---GCGGATATCGAC 948

Qy    335 LysGlyArgValGluGluProThrArgCysGluLysProGluCysAlaAlaArgAsnAla 354
Db    949 CGAGGAAGAATCAGCGAGCCAGAGCGATGTCCCCGAGACGTCTGTGGTTCAACTGGCACT 1008

Qy    355 MetSerLeuIleHisAsnArgCysThrPheAlaAsnLysGlnIleValArgLeuGlnGlu 374
Db   1009 ATGTCTCTTATTCATAACCGCTCCGAATTCACCGACAAACAAGTCATCCGTCTTCAGGAG 1068

Qy    375 ThrProAspAlaIleProGluGlyGluThrProHisThrValSerMetCysLeuTyrAsn 394
Db   1069 ACTCCCGATGCTGTACCCGATGGGCAGACTCCACATACTGTATCTCTTTGCGTCTATGAT 1128

Qy    395 ThrMetValAspAlaValLysProGlyAspArgIleGluValThrGlyValPheLysAla 414
Db   1129 GAACTCGTTGACTTGGTCAAGCCTGGTGACCGAGTGATCATCACTGGTATCTTCCGATCA 1188

Qy    415 MetAlaValArgValGlyProAsnGlnArgThrLeuArgAlaLeuTyrLysThrTyrIle 434
Db   1189 ATCCCAGTTCGTGTCAACCCTCGCCAAAGGAGTATCAAATCTTTGTACAAGACCTATCTC 1248

Qy    435 AspCysValHisValLysLysSerAspArgGlyArgLeuGlnThrGluAspProMetGlu 454
Db   1249 GATGTCGTCCACGTCAAGCGTACCAACACTGCCCGAATGGGCTTT---GACCCTTCTACT 1305

Qy    455 MetAspLysGluAsnAspMetTyrAlaGlyTyrHisGluSerAspThrSerGluAlaAla 474
Db   1306 CGTGCGGGCGAGAGCAAGCCA---CCGGGTGACCTTGGGTCGCCTGTCAGATCAGCCGCT 1362

Qy    475 AsnGluAlaLysIleGlnLysLeuLysGluLeuSerLysLeuProGlyIleTyrAspArg 494
Db   1363 GCGGAGATGGAA---CAACGCATTATCGAGCTCTCCAACCACCCTGATCTTTACAACATA 1419

Qy    495 LeuSerArgSerLeuAlaProSerIleTrpGluLeuGluAspIleLysLysGlyLeuLeu 514
Db   1420 CTTGCCTCATCTTTGGCGCCCTCCATCTACGAGCTCGAAGATGTTAAGAAGGGCATCCTC 1479

Qy    515 CysGlnLeuPheGlyGlyLysAlaLysLysIleProSerGlyAlaSer------------ 530
Db   1480 CTCCAGCTTTTCGGCGGTACCAACAAGTCTATCGCCCGTGGTGGCGGTGGGGGCGGTCCT 1539

Qy    531 ---PheArgGlyAspIleAsnValLeuLeuValGlyAspProGlyThrSerLysSerGln 549
Db   1540 AGATACAGAGGTGATATCAACGTGTTGATGGTTGGTGATCCTGGTACTAGTAAATCTCAG 1599

Qy    550 LeuLeuGlnTyrValHisLysIleAlaProArgGlyIleTyrThrSerGlyArgGlySer 569
Db   1600 ATTCTGCAGTATGTGCATAAGATTGCTCCTCGAGGTGTCTACACCTCCGGTAAGGGAAGT 1659

Qy    570 SerAlaValGlyLeuThrAlaTyrValThrLysAspProGluThrArgGluThrValLeu 589
Db   1660 TCAGCGGTTGGTTTGACGGCCTATGTAACGAGAGATCCCGATTCCAAGCAGCTCGTTTTG 1719

Qy    590 GluSerGlyAlaLeuValLeuSerAspArgGlyIleCysCysIleAspGluPheAspLys 609
Db   1720 GAAAGTGGTGCATTGGTTTTGTCTGACGGCGGAGTTTGCTGTATCGATGAGTTTGACAAG 1779

Qy    610 MetSerAspAsnAlaArgSerMetLeuHisGluValMetGluGlnGlnThrValSerVal 629
Db   1780 ATGAGTGATGCCACTCGAAGTGTCTTGCACGAAGTAATGGAGCAACAAACGGTTTCTATT 1839

Qy    630 AlaLysGlyGlyIleIleAlaSerLeuAsnAlaArgThrSerValLeuAlaCysAlaAsn 649
```

```
                 ||||||   |||||||||   :::|||||||||||||||||:::||||||   ||||||
Db       1840 GCCAAGGCAGGTATCATCACCACTCTCAATGCTCGTACATCCATTCTCGCCGCCGCGAAC 1899

Qy        650 ProSerGlySerArgTyrAsnAlaArgLeuSerValIleAspAsnIleGlnLeuProPro 669
              |||      |||||||||:::        |||    :::      ||||||   |||||||||
Db       1900 CCAATCAACTCTCGATATGACCCTAACCTCCCTATTCCTGCCAACATCGATCTCCCTCCT 1959

Qy        670 ThrLeuLeuSerArgPheAspLeuIleTyrLeuMetLeuAspLysProAspGluGlnAsn 689
              ||||||:::||||||||||||||||:::||||||:::|||||||||   ||||||   |||
Db       1960 ACATTGATCTCAAGATTCGACTTGCTGTACTTGGTGTTGGACAAGGTGGATGAGGTGAAC 2019

Qy        690 AspArgArgLeuAlaArgHisLeuValAlaLeuHisTyrGluAsnTyrGluValSerLys 709
              ||||||:::||||||:::|||||||||      |||:::      :::   |||
Db       2020 GACAGAAAGTTGGCAAAGCATTTGGTGGGACTCTATTTGTCAGATGTAGAGGATCAGCCT 2079

Qy        710 GlnAsp---AlaLeuAspLeuGlnThrLeuThrAlaTyrIleThrTyrAlaArgGlnHis 728
                 |||      :::   |||||||||||||||:::||||||  ||||||||||
Db       2080 GCCGACAATATCATTCCTCTTCAGACCTTAACGTCCTACATTACATATGCCCGCTCCAAA 2139

Qy        729 ValHisProThrLeuSerAspGluAlaAlaGluAspLeuIleAsnGlyTyrValGluMet 748
              :::||||||   |||::::::   |||:::|||   |||:::      ||||||||||||
Db       2140 ATCCACCCCGTCCTCACCGAAGGAGCTTCTGAAGCGCTTGTTCAGGCCTATGTTGAGATG 2199

Qy        749 ArgGlnLysGlyAsnPheProGlySerSerLysLysValIleThrAlaThrProArgGln 768
              |||:::   |||            :::   :::|||   ||||||||||||   ||||||
Db       2200 CGCAAGGCCGGTATGGACTCTCGTACTCAGGAAAAGAGGATCACTGCTACCACTAGGCAG 2259

Qy        769 LeuGluSerMetIleArgIleSerGluAlaLeuAlaArgMetArgPheSerGluValVal 788
              ||||||||||||||||||:::      ||||||   ||||||||||||   |||:::   |||
Db       2260 TTGGAAAGTATGATCAGATTGGGCGAAGCGCATGCGAGGATGAGATTAAGTGACAGGGTT 2319

Qy        789 GluLysValAspAlaAlaGluAlaValArgLeuLeuAspValAlaLeuGlnGlnSerAla 808
              |||:::   |||                          ||||||||:::::|||||||
Db       2320 GAGGAAGAGGACATTAGGGAGGCTGTTAGGTTGATCAAGAGTGCCCTGAGGGAGAGTGCT 2379

Qy        809 ThrAspHisAlaThrGlyThrIleAspMetAspLeuIleThrThrGlyValSerAlaSer 828
              ||||||      ||||||   ||||||:::|||||||||   ||||||             :::
Db       2380 ACCGATCCCCTTACTGGTCAAATTGATCTTGACCTCATCAACACCGGTGCCGGCCAAACC 2439

Qy        829 GluArgIleArgArgAlaAsnLeuLeuAlaAlaLeuArgGluLeuIleAlaAspLysIle 848
                 |||      ||||||:::|||            :::   :::|||:::   :::|||
Db       2440 ATGCGTCGCGCCCGAGCCGATCTTAAACGTGAAGTTATCAAACTCGTCGTCGAAAAG--- 2496

Qy        849 SerProGlySerSerSerGlyLeuLysThrSerGlnLeuLeuGluAspIleArgSerGln 868
                             |||   |||::::::   :::   ::::::::::::::      |||
Db       2497 ------GCCCGATCTCAGGGTATTCGCTGGGCAGCAGTCATTGATGAGCTCAACAAACAA 2550

Qy        869 SerSerValAspValSerLeuGlnAspIleLysAsnAlaLeuGlySerLeuGlnGlyGlu 888
              |||||||||   |||                   :::      |||:::   |||
Db       2551 TCCAGTGTTCCCGTGGATCATGCACAGTTCGCAGAAATTGTTAGGGAGCTGGAAGAAGAG 2610

Qy        889 GlyPheLeuThrValHisGlyAspIleValLysArgVal 901
                    :::   |||   |||:::      :::|||:::
Db       2611 AGCATAGTCAAGGTCATGGGTGAAAGGGAAAGAAGGATT 2649
```

RESULT 11
US-10-369-493-46198
; Sequence 46198, Application US/10369493
; Publication No. US20030233675A1
; GENERAL INFORMATION:
; APPLICANT: Cao, Yongwei
; APPLICANT: Hinkle, Gregory J.
; APPLICANT: Slater, Steven C.
; APPLICANT: Goldman, Barry S.
; APPLICANT: Chen, Xianfeng
; TITLE OF INVENTION: EXPRESSION OF MICROBIAL PROTEINS IN PLANTS FOR PRODUCTION OF
; TITLE OF INVENTION: PLANTS WITH IMPROVED PROPERTIES
; FILE REFERENCE: 38-10(52052)B
; CURRENT APPLICATION NUMBER: US/10/369,493
; CURRENT FILING DATE: 2003-02-28
; PRIOR APPLICATION NUMBER: US 60/360,039
; PRIOR FILING DATE: 2002-02-21
; NUMBER OF SEQ ID NOS: 47374
; SEQ ID NO 46198
;   LENGTH: 2802
;   TYPE: DNA
;   ORGANISM: Saccharomyces cerevisiae
US-10-369-493-46198

Alignment Scores:

Pred. No.:	3.95e-147	Length:	2802
Score:	1833.00	Matches:	393
Percent Similarity:	59.36%	Conservative:	178
Best Local Similarity:	40.85%	Mismatches:	281
Query Match:	39.77%	Indels:	110
DB:	6	Gaps:	14

US-10-768-511-8 (1-901) x US-10-369-493-46198 (1-2802)

```
Qy          2 GluAsnAsnAspAlaLeuAspIleGlyAlaValSerSerProTyrProSerGlnSerGlu 21
              |||:::|||::::::                   |||||||||             :::
Db         31 GAGGATAATAACTCC------------------AGCTCCCCGTTGTGCCTAATCCTGAT 72

Qy         22 GlyValSerThrProLeuProGlnValThrSerProSerPheAspAsnAlaAlaSerPro 41
                 |||                 ||||||::::::||||||:::         :::::::|||
Db         73 TCTGTTCCA---------CCACAGCTTTCTTCCCCAGCTCTATTTATAGCTCCTCTTCA 123

Qy         42 ValAlaGly-------------------------------------------------- 44
                    |||
Db        124 TCACAAGGTGATATCTATGGTCGCAACAATAGCCAGAACTTAAGTCAGGGAGAGGGAAAC 183

Qy         45 ---ArgArgAlaValArgGlnThrPro---------ThrSerAlaValArgArgArgGly 60
                 |||   |||:::      :::|||            |||:::   :::|||:::
Db        184 ATCAGAGCTGCTATAGGTTCTTCTCCACTAAATTTTCCATCTTCTTCCCAAAGACAAAAT 243

Qy         61 ArgGluThrAspSerAlaArgArgArgArgSerArgSerArgSerLeuGlyAsnSerVal 80
                 :::         ::::::   |||:::   |||   ||||||         ::::::
Db        244 TCCGATGTTTTCCAATCTCAAGGCAGACAGGGCAGAATTCGTTCTTCTGCGAGCGCTTCT 303

Qy         81 TyrSerSerProTyrAspAlaGlyThrProGlyThrProGlyThrProValAlaThrPro 100
                       |||   |||   :::                        |||   ::::::
Db        304 GGAAGGTCTAGATATCATTCTGATCTGAGAAGTGATAGAGCACTGCCTACTTCTTCC--- 360

Qy        101 ValTyrAlaThrProValGlyThrProMetGlyThrProSerPheHisArgGlyThrPro 120

Db        360 ------------------------------------------------------------ 360

Qy        121 GlnTyrLysGlnArgSerGluLeuGlySerGlnGlyLysProLeuHisArgArgArgArg 140
                                |||   ||||||   |||:::               ||||||
Db        361 ---------------TCTTCTTTAGGCCGTAATGGTCAAAACCGTGTACACATGCGAAGA 405

Qy        141 SerGlnSerArgGluProGlyHisArgSerProSerArgGluProSerAlaAspGlyArg 160
              :::                                ||||||   |||            |||   |||
Db        406 AATGATATTCATACATCTGATTTATCCTCTCCAAGAAGAATTGTGGATTTTGATACTAGA 465

Qy        161 ------------------------------------------ProSerGluSerAlaGlu 166
                                                          |||||||||::::::|||
Db        466 TCCGGCGTGAATACTTTGGATACTTCTTCTTCTTCTGCTCCTCCATCGGAAGCCAGTGAA 525

Qy        167 ProAspAspThrLeuGlyGlyGluTyrAlaTyrValTrpGlyThrAsnValAsnIlePro 186
              |||                                 :::|||||||||||||||:::|||
Db        526 CCC---------------------TTGAGAATAATTTGGGGTACCAACGTCAGTATCCAG 564

Qy        187 AspValLeuArgAlaIleArgArgPheLeuHisAsnTyrArg------------------ 200
              :::                  |||   ||||||   :::::::
Db        565 GAATGTACAACTAATTTCGTAATTTTTGATGTCTTTCAAGTATAAATTCCGTAAAATA 624

Qy        201 ------------------------SerSerAlaHisAspLeuAsnSerLysTyrIleGln 212
                                      ::::::      |||         ||||||:::
Db        625 TTGGATGAAAGGGAGGAATTCATTAACAATACTACCGACGAAGAACTATACTATATCAAG 684

Qy        213 IleIleGluGluThrValGluArgGluGluAspThrLeuAsnIleAspMetSerAspIle 232
                 :::   |||      |||            ||||||:::|||      ::::::
Db        685 CAGCTTAATGAAATGAGAGAACTTGGTACCTCTAATTTAAACTTGGATGCTAGAAACTTA 744

Qy        233 ------TyrAspHisAspProAspLeuTyrAlaLysIleValArgTyrProLeuAspIle 250
```

```
Db        745 CTTGCTTACAAGCAAACAGAAGACTTATACCATCAACTATTAAATTATCCTCAAGAAGTG 804
Qy        251 IleProLeuLeuAspThrGluCysGlnGluValAlaThrSerLeu--------------- 265
Db        805 ATTTCTATCATGGATCAAACTATCAAGGACTGTATGGTTTCCTTAATAGTAGACAACAAT 864
Qy        266 ------------LeuProThrPheGlu---LysHisIleGluAlaArgProPheAsnLeu 280
Db        865 TTGGATTACGATTTAGATGAAATAGAGACCAAGTTTTATAAAGTAAGGCCTTACAATGTA 924
Qy        281 LysAlaSerValHisMetArgGluLeuAsnProSerAspIleAspLysLeuValSerVal 300
Db        925 GGTTCTTGTAAAGGTATGCGCGAATTGAATCCAAATGATATTGATAAATTGATAAACTTA 984
Qy        301 LysGlyMetValIleArgCysSerSerIleIleProGluIleLysGlyAlaPhePheLys 320
Db        985 AAGGGTCTTGTCCTGAGATCAACTCCGGTTATCCCTGATATGAAGGTAGCGTTTTTCAAA 1044
Qy        321 CysLeuValCysGlyHisSerProProLeuValThrValValLysGlyArgValGluGlu 340
Db       1045 TGCAACGTCTGCGATCATACA---ATGGCAGTGGAAATTGATAGAGGAGTTATACAAGAG 1101
Qy        341 ProThrArgCysGluLysProGluCysAlaAlaArgAsnAlaMetSerLeuIleHisAsn 360
Db       1102 CCCGCTAGGTGTGAACGTATTGATTGTAATGAACCCAATTCCATGTCACTGATTCACAAC 1161
Qy        361 ArgCysThrPheAlaAsnLysGlnIleValArgLeuGlnGluThrProAspAlaIlePro 380
Db       1162 AGGTGTTCATTCGCAGATAAACAAGTCATTAAGTTACAGGAAACTCCAGACTTTGTGCCT 1221
Qy        381 GluGlyGluThrProHisThrValSerMetCysLeuTyrAsnThrMetValAspAlaVal 400
Db       1222 GATGGACAAACGCCTCACTCTATCTCATTATGTGTTTACGATGAATTAGTGGATTCTTGC 1281
Qy        401 LysProGlyAspArgIleGluValThrGlyValPheLysAlaMetAlaValArgValGly 420
Db       1282 AGGGCGGGCGATCGTATTGAGGTGACTGGCACGTTCAGGTCCATCCCCATTAGAGCTAAT 1341
Qy        421 ProAsnGlnArgThrLeuArgAlaLeuTyrLysThrTyrIleAspCysValHisValLys 440
Db       1342 TCCAGGCAACGCGTACTAAAGTCGTTGTATAAAACATACGTCGATGTGGTCCACGTTAAA 1401
Qy        441 LysSerAspArgGlyArgLeuGlnThrGluAspProMetGluMetAspLysGluAsnAsp 460
Db       1402 AAAGTTTCAGATAAGAGGTTAGACGTCGATACTTCTACTATTGAACAAGAATTAATGCAG 1461
Qy        461 MetTyrAlaGlyTyrHisGluSerAspThrSerGluAlaAlaAsnGluAlaLysIleGln 480
Db       1462 AACAAGGTAGATCATAACGAGGTCGAAGAAGTAAGACAAATTACTGATCAGGATTTAGCA 1521
Qy        481 LysLeuLysGluLeuSerLysLeuProGlyIleTyrAspArgLeuSerArgSerLeuAla 500
Db       1522 AAAATTCGCGAGGTTGCGGCAAGAGAAGATTTATACAGTTTATTAGCCCGCTCTATTGCC 1581
Qy        501 ProSerIleTrpGluLeuGluAspIleLysLysGlyLeuLeuCysGlnLeuPheGlyGly 520
Db       1582 CCAAGTATTTACGAGCTAGAAGATGTCAAGAAGGGTATATTACTTCAGCTATTTGGCGGC 1641
Qy        521 LysAlaLysLysIleProSerGlyAlaSerPheArgGlyAspIleAsnValLeuLeuVal 540
Db       1642 ACGAATAAAACCTTTACCAAGGGTGGCCGTTATAGAGGTGACATAAATATTTTACTTTGT 1701
Qy        541 GlyAspProGlyThrSerLysSerGlnLeuLeuGlnTyrValHisLysIleAlaProArg 560
Db       1702 GGGGATCCTTCTACTTCCAAATCGCAAATTTTGCAATACGTTCACAAAATTACTCCTCGT 1761
Qy        561 GlyIleTyrThrSerGlyArgGlySerSerAlaValGlyLeuThrAlaTyrValThrLys 580
Db       1762 GGTGTGTATACTTCGGGTAAAGGTTCATCTGCCGTTGGTTTAACTGCTTATATTACAAGG 1821
Qy        581 AspProGluThrArgGluThrValLeuGluSerGlyAlaLeuValLeuSerAspArgGly 600
Db       1822 GATGTCGACACAAAACAACTTGTTTTGGAAAGTGGTGCATTAGTATTGTCTGATGGAGGT 1881
Qy        601 IleCysCysIleAspGluPheAspLysMetSerAspAsnAlaArgSerMetLeuHisGlu 620
Db       1882 GTTTGTTGTATTGACGAGTTTGATAAAATGAGTGATTCTACAAGGTCCGTCTTGCACGAA 1941
Qy        621 ValMetGluGlnGlnThrValSerValAlaLysGlyGlyIleIleAlaSerLeuAsnAla 640
Db       1942 GTCATGGAACAGCAGACTATTTCAATCGCAAAAGCGGGAATTATCACAACACTTAATGCC 2001
Qy        641 ArgThrSerValLeuAlaCysAlaAsnProSerGlySerArgTyrAsnAlaArgLeuSer 660
Db       2002 AGAAGTTCTATTTTGGCCAGTGCTAACCCAATTGGTTCACGCTACAACCCCAATTTGCCT 2061
Qy        661 ValIleAspAsnIleGlnLeuProProThrLeuLeuSerArgPheAspLeuIleTyrLeu 680
Db       2062 GTGACTGAAAATATTGATCTACCGCCCCCACTACTTTCGAGATTCGATCTGGTCTATCTT 2121
Qy        681 MetLeuAspLysProAspGluGlnAsnAspArgArgLeuAlaArgHisLeuValAlaLeu 700
Db       2122 GTCCTTGATAAGGTTGATGAGAAAAATGACAGAGAACTAGCCAAACACTTAACAAATCTT 2181
Qy        701 HisTyrGluAsn------TyrGluValSerLysGlnAspAlaLeuAspLeuGlnThrLeu 718
Db       2182 TACCTGGAAGATAAGCCCGAACATATTTCTCAAGACGACGTGCTACCAGTTGAATTTTTA 2241
Qy        719 ThrAlaTyrIleThrTyrAlaArgGlnHisValHisProThrLeuSerAspGluAlaAla 738
Db       2242 ACGATGTATATTAGTTATGCAAAGGAGCACATACACCCAATAATCACCGAGGCCGCTAAG 2301
Qy        739 GluAspLeuIleAsnGlyTyrValGluMetArgGlnLysGlyAsnPheProGlySerSer 758
Db       2302 ACTGAGCTTGTTCGTGCTTATGTAGGAATGAGAAAGATGGGTGACGATTCGAGATCCGAT 2361
Qy        759 LysLysValIleThrAlaThrProArgGlnLeuGluSerMetIleArgIleSerGluAla 778
Db       2362 GAGAAGAGAATCACAGCTACCACAAGACAACTTGAAAGTATGATTCGTTTGGCTGAGGCG 2421
Qy        779 LeuAlaArgMetArgPheSerGluValValGluLysValAspAlaAlaGluAlaValArg 798
Db       2422 CACGCCAAAATGAAATTGAAAAACGTCGTAGAGCTGGAGGATGTTCAAGAAGCCGTTAGA 2481
Qy        799 LeuLeuAspValAlaLeuGlnGlnSerAlaThrAspHisAlaThrGlyThrIleAspMet 818
Db       2482 TTAATTAGATCAGCCATAAAAGATTATGCAACAGACCCTAAAACCGGTAAAATCGACATG 2541
Qy        819 AspLeuIleThrThrGlyValSerAlaSerGluArgIleArgArgAlaAsnLeuLeuAla 838
Db       2542 AATTTAGTTCAAACAGGTAAATCAGTTATTCAGAGAAAACTACAGGAGGATTTGTCAAGG 2601
Qy        839 AlaLeuArgGluLeuIleAlaAspLysIleSerProGlySerSerSerGlyLeuLysThr 858
Db       2602 GAAATTATGAATGTTTTGAAGGATCAG------------GCATCAGACTCAATGTCATTC 2649
Qy        859 SerGlnLeuLeuGluAspIleArgSerGlnSerSerValAspValSerLeuGlnAspIle 878
Db       2650 AATGAGCTTATAAAACAAATCAATGAACACTCTCAAGATAGAGTTGAGTCTTCTGATATC 2709
Qy        879 LysAsnAlaLeuGlySerLeuGlnGlyGluGlyPheLeuThrValHisGlyAspIleVal 898
Db       2710 CAGGAAGCCTTGTCAAGATTGCAACAAGAGGACAAGGTCATTGTCCTTGGCGAGGGTGTA 2769
Qy        899 LysArg 900
Db       2770 AGGAGA 2775
```

RESULT 12
US-10-369-493-27488
; Sequence 27488, Application US/10369493
; Publication No. US20030233675A1
; GENERAL INFORMATION:
; APPLICANT: Cao, Yongwei
; APPLICANT: Hinkle, Gregory J.

```
; APPLICANT:  Slater, Steven C.
; APPLICANT:  Goldman, Barry S.
; APPLICANT:  Chen, Xianfeng
; TITLE OF INVENTION: EXPRESSION OF MICROBIAL PROTEINS IN PLANTS FOR PRODUCTION OF
; TITLE OF INVENTION:  PLANTS WITH IMPROVED PROPERTIES
; FILE REFERENCE: 38-10(52052)B
; CURRENT APPLICATION NUMBER: US/10/369,493
; CURRENT FILING DATE:  2003-02-28
; PRIOR APPLICATION NUMBER: US 60/360,039
; PRIOR FILING DATE: 2002-02-21
; NUMBER OF SEQ ID NOS: 47374
; SEQ ID NO 27488
;   LENGTH: 2881
;   TYPE: DNA
;   ORGANISM: Neurospora crassa
US-10-369-493-27488

Alignment Scores:
Pred. No.:              6.57e-146     Length:        2881
Score:                  1819.00       Matches:       419
Percent Similarity:     58.17%        Conservative:  161
Best Local Similarity:  42.03%        Mismatches:    249
Query Match:            39.47%        Indels:        169
DB:                     6             Gaps:          24

US-10-768-511-8 (1-901) x US-10-369-493-27488 (1-2881)

Qy       24 SerThrProLeuProGlnValThrSerProSerPheAsp--------------------- 36
Db        1 AGCTCGCCTTTGCGCCAGCAGACAGAGACCCAGAGCGATGCTGACCGAACTCCTCGCGCA 60

Qy       37 AsnAlaAlaSerProValAlaGlyArgArgAlaValArgGlnThr--------------- 51
Db       61 AACGGACGGAGCCAGTTAATTGGGGTATGGCTGTCAACCAAAAAAAAAAAAAAAAAGGAC 120

Qy       52 ---------------ProThrSerAlaValArgArgArgGlyArgGluThrAspSerAla 66
Db      121 TACACTGAATGCTGACCACACTCCGCA------------------------GACTCCTCT 156

Qy       67 ArgArgArgArgSerArgSerArgSerLeuGlyAsn-SerVal.TyrSerSerProTyr-- 85
Db      157 CCCATCCGCTATGCCTCCAGTTCTAGCCCGGGCCGTCAGCTTACTCAACAGTCCGATCTT 216

Qy       86 -------------AspAlaGlyThrProGlyThrProGlyThrProValAlaThrProVa 101
Db      217 CGTAGCGAAAGCAGTCAGCTCTTCGTCAGCTCGCAAAGGTCCGTTGCCGGTCGATCCCGT 276

Qy      101 lTyrAlaThrProValGlyThrProMetGlyThrProSerPheHisArgGlyThrProGl 121
Db      277 CGAGGCGACATCAACGGGGAC-CCTCTTCGTACTCCCGCC-------------------CA 317

Qy      121 nTyrLysGlnArgSerGluLeuGlySerGlnGlyLysProLeuHisArgArgArgArgSe 141
Db      318 GATTCCCCGCCGAATCATCCTCGACGACGCCGGTCGTGTGATCCGC-------------- 363

Qy      141 rGlnSerArgGluProGlyHisArgSerProSerArgGluProSerAlaAspGlyArgPr 161
Db      364 ----------------------------------GATGCCCCTGGCTCCGATGCCAACTC 389

Qy      161 oSerGluSerAlaGluProAsp------------AspThrLeuGlyGlyGlu---TyrAl 176
Db      390 GTTTGTCACCAATAACCCCAATACCTCTGAGGCAGATGCCCTTGGTGGACAAAGCCAAGG 449

Qy      176 aTyrValTrpGlyThrAsnValAsnIleProAspValLeuArgAlaIleArgArgPheLe 196
Db      450 TCTTGTTTGGGGTACAACAATTTCTCTTGACGATTCGTTCTCCGCGTTCAAGGACTTTCT 509

Qy      196 uHisAsn------------TyrArgSerSerAlaHis----------------------- 204
Db      510 CAGGAACTTCACCAGGAAATACCGGATGTGGGCAGATGGCGCGGACGAGGCCGAGACTAT 569

Qy      205 ----------AspLeuAsnSerLys---TyrIleGlnIleIleGluGluThrValGluAr 220
Db      570 TGGCCATCCGGATGCGGATTCGAAGCCATATTGGGAGGCACTGGAGAATATGCTGCTGCT 629

Qy      220 gGluGluAspThrLeuAsnIleAspMetSerAspIle------TyrAspHisAspProAs 238
Db      630 TGGTACCAACAAGCTGTATCTGGATCTCCGCGACCTCAAGTCCTACCCGCGCACCTTGAA 689

Qy      238 pLeuTyrAlaLysIleValArgTyrProLeuAspIleIleProLeuLeuAspThr----- 256
Db      690 GTTATGGCATCAGGCCCAACACTATCCTACCGAAATCATCCCGGTCATGGATCAATGCGT 749

Qy      257 ----GluCys------------------GlnGluValAlaThr----------------- 263
Db      750 ACATGACTGTATGATGGAGTTGGCGCAAAAGGAAATGGCCAGCCAAAGAGCATCCCAAAA 809

Qy      263 ------------------------------------------------------------ 263
Db      810 CTCTAGGACCGCTCCAGGTGCCTCACAAAGCTCGGAGCCGAACTTCCCAAGCTCTGAAAG 869

Qy      264 --------------------------------SerLeuLeuProThrPheGluLysHi 272
Db      870 GAGCGAGGAACCGCCCACCCCGCGGCCTGCCCAAACAGCGGCGCCCACAATTGAGGACCA 929

Qy      272 sIleGlu---------------AlaArgProPheAsnLeuLysAlaSerValHisMetAr 287
Db      930 AGTCAGCCAGATGGCGTATGTGGTTCGCCCGTGGGGTCTTGACAAGATCACCAACCTCCG 989

Qy      287 gGluLeuAsnProSerAspIleAspLysLeuValSerValLysGlyMetValIleArgCy 307
Db      990 TGACCTCAACCCATCTGACATGGACAAGCTTGTGTCCATCAAGGGTCTTGTCATCAGAAC 1049

Qy      307 sSerSerIleIleProGluIleLysGlyAlaPhePheLysCysLeuValCysGlyHisSe 327
Db     1050 CACGCCGGTCATCCCCGACATGAAGGATGCATTTTTCAAGTGCAGTGTTTGTGGTCATTC 1109

Qy      327 rProProLeuValThrValVal------LysGlyArgValGluGluProThrArgCysGl 345
Db     1110 T---------ATAACCGTTCAGTTAGATCGTGGCAAGATTCGTGAGCCTACAGAGTGTCC 1160

Qy      345 uLysProGluCysAlaAlaArgAsnAlaMetSerLeuIleHisAsnArgCysThrPheAl 365
Db     1161 TCGGGCTCGCTGCGCCTCCAAGAACTCAATGCAGATCATCCACAACAGATGTGCTTTCGA 1220

Qy      365 aAsnLysGlnIleValArgLeuGlnGluThrProAspAlaIleProGluGlyGluThrPr 385
Db     1221 AGACAAGCAAGTCATCAAGCTGCAAGAAACTCCGGATAACGTGCCTGCTGGTCAGACACC 1280

Qy      385 oHisThrValSerMetCysLeuTyrAsnThrMetValAspAlaValLysProGlyAspAr 405
Db     1281 CCACTCGGTATCCGTTTGCGTCTATAACGAGCTTGTGGACTTTTGCAAGGCCGGTGACAG 1340

Qy      405 gIleGluValThrGlyValPheLysAlaMetAlaValArgValGlyProAsnGlnArgTh 425
Db     1341 AGTTGAGCTCACTGGTATCTTTAAAGTCACCCCCGTTCGTGTCAACCCGCGCATGAGGAC 1400

Qy      425 rLeuArgAlaLeuTyrLysThrTyrIleAspCysValHisValLysLysSerAspArgGl 445
Db     1401 GGTCAAGAGCGTGCACAAGACATATGTTGATGTTGTCCATGTACAAAAGGTCGACCGCAA 1460

Qy      445 yArgLeuGlnThrGluAspProMetGluMetAsp------LysGluAsnAspMetTyrAl 463
Db     1461 GAGGATGGGTTCT---GATCCCTCGACCCTCGACCTTGCCGAGGAGGAGGAAGCCCATGC 1517

Qy      463 aGlyTyrHisGluSerAspThrSerGluAlaAlaAsnGluAlaLysIleGlnLysLeuLy 483
Db     1518 TAATGGCCAGAGCATGGACGAAGTCCGAAAGGTCTCTCCTGATGAGGAAGAGAGGATCAA 1577

Qy      483 sGluLeuSerLysLeuProGlyIleTyrAspArgLeuSerArgSerLeuAlaProSerIl 503
Db     1578 GGAAACTGCCGCTCGACCGGATATCTACGATCTCCTTTCACGTTCTTTGGCTCCATCAAT 1637

Qy      503 eTrpGluLeuGluAspIleLysLysGlyLeuLeuCysGlnLeuPheGlyGlyLysAlaLy 523
Db     1638 TTACGAGATGGATGATGTGAAGAAGGGTATTCTTCTTCAACTGTTCGGAGGCACGAACAA 1697
```

```
Qy      523 sLysIleProSerGlyAlaSer------PheArgGlyAspIleAsnValLeuLeuValGl 541
Db     1698 GACCTTTGAGAAGGGTGGAAGCCCCAAATACCGTGGAGACATCAACGTTCTCCTCTGTGG 1757

Qy      541 yAspProGlyThrSerLysSerGlnLeuLeuGlnTyrValHisLysIleAlaProArgGl 561
Db     1758 TGACCCATCCACTTCCAAGTCCCAGCTTCTCTCCTATGTTCACAGAATCGCCCCAAGAGG 1817

Qy      561 yIleTyrThrSerGlyArgGlySerSerAlaValGlyLeuThrAlaTyrValThrLysAs 581
Db     1818 TGTGTACACGAGTGGTAAAGGTTCTTCTGCCGTAGGTCTTACCGCCTATGTTACCCGCGA 1877

Qy      581 pProGluThrArgGluThrValLeuGluSerGlyAlaLeuValLeuSerAspArgGlyIl 601
Db     1878 TCCCGAGTCCCGCCAACTCGTTCTAGAGTCCGGTGCTCTCGTCTTATCCGACGGCGGTGT 1937

Qy      601 eCysCysIleAspGluPheAspLysMetSerAspAsnAlaArgSerMetLeuHisGluVa 621
Db     1938 CTGTTGTATCGACGAGTTCGACAAGATGAACGAATCGACCCGCTCCGTTCTTCACGAAGT 1997

Qy      621 lMetGluGlnGlnThrValSerValAlaLysGlyGlyIleIleAlaSerLeuAsnAlaAr 641
Db     1998 CATGGAACAACAGACCGTGTCCGTCGCCAAGGCCGGCATCATCACCACTCTCAATGCCCG 2057

Qy      641 gThrSerValLeuAlaCysAlaAsnProSerGlySerArgTyrAsnAlaArgLeuSerVa 661
Db     2058 AACATCAATCCTGGCGTCTGCCAACCCCATTGGTAGTAGATACAACCCTGATCTATCAGT 2117

Qy      661 lIleAspAsnIleGlnLeuProProThrLeuLeuSerArgPheAspLeuIleTyrLeuMe 681
Db     2118 GCCGCAAAATATTGACCTCCCGCCAACCCTTCTATCCCGTTTCGATCTGGTCTACCTCAT 2177

Qy      681 tLeuAspLysProAspGluGlnAsnAspArgArgLeuAlaArgHisLeuValAlaLeuHi 701
Db     2178 CCTCGATCGTGTTGATGAGAAGAATGATCAGCGTCTGGCCCGCCATTTGCTCTCCATGTA 2237

Qy      701 sTyrGluAsnTyrGluValSerLysGlnAspAlaLeuAspLeu---------------- 715
Db     2238 TCTCGAAGACAAGCCCGAGAGTGCCCAGCAGGCCAACGACGTCCTTGTGAGTTTTTCCCC 2297

Qy      715 ------------------------------------------------------------ 715
Db     2298 GTCTGTCAATCCACACGAAGATGAACCCCAAACTAACCAACTCCCCCTCAAAAAAACAGC 2357

Qy      716 ----GlnThr-LeuThrAlaTyrIleThrTyrAlaArgGlnHisValHisProThrLeuS 734
Db     2358 CCGTCGAGTTTCTAACCTCCTACATCTCCTACGCCCGCTCGCACATCCACCCGGCCCTCA 2417

Qy      734 erAspGluAlaAlaGluAspLeuIleAsnGlyTyrValGluMetArgGlnLysGlyAsnP 754
Db     2418 CGCCCGAAGCCGGTCGCGAGCTCGTCGATGCCTACGTCGAGATGCGCAAGCTGGGCCAAG 2477

Qy      754 heProGlySerSerLysLysValIleThrAlaThrProArgGlnLeuGluSerMetIleA 774
Db     2478 ACGTGCGCGCCGCCGAGAAGCGCATTACGGCCACCACACGTCAGCTCGAGTCCATGATTC 2537

Qy      774 rgIleSerGluAlaLeuAlaArgMetArgPheSerGluValValGluLysValAspAlaA 794
Db     2538 GCTTAGCCGAGGCGCACGCAAAGATGCGTTTATCTCAAACCGTCACGCGCGACGACGTCC 2597

Qy      794 laGluAlaValArgLeuLeuAspValAlaLeuGlnGlnSerAlaThrAspHisAlaThrG 814
Db     2598 GCGAGGCCGTGCGCCTCATCAAGTCGGCGCTCAAGACCGCCGCCACCGAC---AGCCAGG 2654

Qy      814 lyThrIleAspMetAspLeuIleThrThrGlyValSerAlaSerGluArgIleArgArgA 834
Db     2655 GCCGCATCGACATGAGCTTGTTGACAGAGGGCACGAGCGCGGCGGAGCGCCAGCGCAAGG 2714

Qy      834 laAsnLeuLeuAlaAlaLeuArgGluLeuIleAlaAspLysIleSerProGlySerSerS 854
Db     2715 CCGACATGAAGGATGCGGTTATCCGCCTGCTGGATGAGATGACCAGCGGCGGACAGGTA- 2773

Qy      854 erGlyLeuLysThrSerGlnLeuLeuGluAspIleArgSerGlnSerSerValAspValS 874
Db     2774 -----GTTAGGTACTCGGAGGTGGCGAGACGGTTGGGCGAGGGGGCGGGTGTGCAGGTTG 2828

Qy      874 erLeuGlnAspIleLysAsnAlaLeuGlySerLeuGlnGlyGluGly 889
Db     2829 AGCCGGCTGAGTTTGCGGAGGTTATGAGGGCTCTGGAGATGGAGGGC 2875
```

RESULT 13
US-09-954-456-568
; Sequence 568, Application US/09954456
; Patent No. US20020115057A1
; GENERAL INFORMATION:
; APPLICANT: Young, Paul
; TITLE OF INVENTION: Process for Identifying Anti-Cancer Therapeutic Agents Using Can
; TITLE OF INVENTION: Sets
; FILE REFERENCE: 689290-76
; CURRENT APPLICATION NUMBER: US/09/954,456
; CURRENT FILING DATE: 2001-09-18
; PRIOR APPLICATION NUMBER: US/60/233,617
; PRIOR FILING DATE: 2000-09-18
; PRIOR APPLICATION NUMBER: US/60/234,052
; PRIOR FILING DATE: 2000-09-20
; PRIOR APPLICATION NUMBER: US/60/234,923
; PRIOR FILING DATE: 2000-09-25
; PRIOR APPLICATION NUMBER: US/60/235,134
; PRIOR FILING DATE: 2000-09-25
; PRIOR APPLICATION NUMBER: US/60/235,637
; PRIOR FILING DATE: 2000-09-26
; PRIOR APPLICATION NUMBER: US/60/235,638
; PRIOR FILING DATE: 2000-09-26
; PRIOR APPLICATION NUMBER: US/60/235,711
; PRIOR FILING DATE: 2000-09-27
; PRIOR APPLICATION NUMBER: US/60/235,720
; PRIOR FILING DATE: 2000-09-27
; PRIOR APPLICATION NUMBER: US/60/235,840
; PRIOR FILING DATE: 2000-09-27
; PRIOR APPLICATION NUMBER: US/60/235,863
; PRIOR FILING DATE: 2000-09-27
; NUMBER OF SEQ ID NOS: 2276
; SOFTWARE: PatentIn version 3.0
; SEQ ID NO 568
; LENGTH: 3273
; TYPE: DNA
; ORGANISM: Homo sapiens
US-09-954-456-568

Alignment Scores:

Pred. No.:	1.05e-145	Length:	3273
Score:	1817.50	Matches:	406
Percent Similarity:	57.78%	Conservative:	151
Best Local Similarity:	42.12%	Mismatches:	256
Query Match:	39.43%	Indels:	151
DB:	3	Gaps:	18

US-10-768-511-8 (1-901) x US-09-954-456-568 (1-3273)

```
Qy       38 AlaAlaSerProValAlaGlyArgArgAlaValArgGlnThrProThrSerAlaValArg 57
Db       21 GCGACGTCGTCGGCAAGCGGCCGC--------------CTTCCACGTAACGCG---CGC 62

Qy       58 ArgArgGlyArgGluThrAspSerAlaArgArgArgArgSerArg-------------- 72
Db       63 CGGCGGGGGAGGGCGTTGGCGCGGAGCCGACGGGAACGTCCGCGCTGCGGAGCAGGGCAG 122

Qy       73 ---SerArgSerLeuGlyAsnSerValTyrSerSerProTyrAspAlaGlyThrProGly 91
Db      123 GGAAGCCGGGAGGCGGGC---CCGGCCCGAGCTTGTCCTTGTCGCGCAGGTACTCCGAGC 179

Qy       92 ThrProGlyThrProValAlaThrProVal-------------------TyrAlaThr 104
Db      180 ACTATGTCGTCCCCGGCGTCGACCCCGAGCCGCCGCGGCAGCCGGCGTGGAAGGGCCACC 239
```

```
Qy      105 ProValGlyThrPro------------------------------------------- 109
Db      240 CCCGCCCAGACGCCTCGGAGTGAGGATGCCAGGTCATCTCCCTCTCAGAGACGTAGAGGC 299

Qy      109 ------------------------------------------------------------ 109
Db      300 GAGGATTCCACCTCCACGGGGGAGTTGCAGCCGATGCCAACCTCGCCTGGAGTGGACCTG 359

Qy      109 ------------------------------------------------------------ 109
Db      360 CAGAGCACTGCTGCGCAGGACGTGCTGTTTTCCAGCCCTCCCCAAATGCATTCTTCAGCT 419

Qy      110 ------------------------------------MetGlyThrProSerPheHis--- 116
Db      420 ATCCCTCTTGACTTTGATGTTAGTTCACCACTGACATACGGCACTCCCAGCTCTCGGGTA 479

Qy      117 ------------------------ArgGlyThrProGlnTyrLysGlnArgSerGluLeu 128
Db      480 GAGGGAACCCCAAGAAGTGGTGTTAGGGGCACACCT---GTGAGACAGAGGCCTGACCTG 536

Qy      129 GlySerGlnGlyLysProLeuHisArgArgArgArgSerGlnSerArgGluProGlyHis 148
Db      537 GGCTCTGCACAGAAGGGCCTGCAAGTGGATCTGCAGTCT--------------------- 575

Qy      149 ArgSerProSerArgGluProSerAlaAspGlyArgProSerGluSerAlaGluProAsp 168
Db      576 --------------------------GACGGGGCAGCAGCAGAAGATATAGTGGCAAGT 608

Qy      169 AspThrLeuGlyGlyGluTyrAlaTyrValTrpGlyThrAsnValAsnIleProAspVal 188
Db      609 GAGCAGTCTCTAGGCCAAAAACTTGTGATCTGGGAACAGATGTAAATGTGGCAGCATGC 668

Qy      189 LeuArgAlaIleArgArgPheLeuHisAsnTyr------------------------Arg 200
Db      669 AAAGAAAACTTTCAGAGATTTCTTCAGCGTTTTATTGACCCTCTGGCTAAAGAAGAAGAA 728

Qy      201 SerSerAlaHisAspLeuAsnSerLys---TyrIleGlnIleIleGluGluThrValGlu 219
Db      729 AATGTTGGCATAGATATTACTGAACCTCTATACATGCAACGACTTGGGGAGATTAATGTT 788

Qy      220 ArgGluGluAspThrLeuAsnIleAspMetSerAspIleTyrAspHisAspProAspLeu 239
Db      789 ATTGGTGAGCAATTTTTAAATGTGAACTGTGAACACATCAAATCATTTGACAAAAATTTG 848

Qy      240 TyrAlaLysIleValArgTyrProLeuAspIleIleProLeuLeuAspThrGluCysGln 259
Db      849 TACAGACAACTCATCTCTTACCCACAGGAAGTTATTCCAACTTTGACATGGCTGTCAAT 908

Qy      260 GluValAlaThrSerLeuLeuPro------ThrPheGluLysHisIleGluAlaArgPro 277
Db      909 GAAATCTTCTTTGACCGTTACCCTGACTCAATCTTAGAACATCAGATTCAAGTAAGACCA 968

Qy      278 PheAsnLeuLysAlaSerValHisMetArgGluLeuAsnProSerAspIleAspLysLeu 297
Db      969 TTCAACGCATTGAAGACTAAGAATATGAGAAACCTGAATCCAGAAGACATTGACCAGCTC 1028

Qy      298 ValSerValLysGlyMetValIleArgCysSerSerIleIleProGluIleLysGlyAla 317
Db     1029 ATCACCATCAGCGGCATGGTGATCAGGACATCCCAGCTGATTCCCGAGATGCAGGAGGCC 1088

Qy      318 PhePheLysCysLeuValCysGlyHisSerProProLeuValThrValValLysGlyArg 337
Db     1089 TTCTTCCAGTGCCAAGTGTGTGCCCACACG---ACCCGGGTGGAGATGGACCGCGGCCGC 1145

Qy      338 ValGluGluProThrArgCysGluLysProGluCysAlaAlaArgAsnAlaMetSerLeu 357
Db     1146 ATTGCAGAGCCCAGTGTGTGCGGGCGC------TGCCACACCACCCACAGCATGGCACTC 1199

Qy      358 IleHisAsnArgCysThrPheAlaAsnLysGlnIleValArgLeuGlnGluThrProAsp 377
Db     1200 ATCCACAACCGCTCCCTCTTCTCTGACAAGCAGATGATCAAGCTTCAGGAGTCTCCGGAA 1259

Qy      378 AlaIleProGluGlyGluThrProHisThrValSerMetCysLeuTyrAsnThrMetVal 397
Db     1260 GACATGCCTGCAGGGCAGACACCACACACAGTTATCCTGTTTGCTCACAATGATCTCGTT 1319

Qy      398 AspAlaValLysProGlyAspArgIleGluValThrGlyValPheLysAlaMetAlaVal 417
Db     1320 GACAAGGTCCAGCCTGGGGACAGAGTGAATGTTACAGGCATCTATCGAGCTGTGCCTATT 1379

Qy      418 ArgValGlyProAsnGlnArgThrLeuArgAlaLeuTyrLysThrTyrIleAspCysVal 437
Db     1380 CGAGTCAATCCAAGAGTGAGTAATGTGAAGTCTGTCTACAAAACCCACATTGATGTCATT 1439

Qy      438 HisValLysLysSerAspArgGlyArgLeuGlnThrGluAspProMetGluMetAspLys 457
Db     1440 CATTATCGGAAAACGGATGCAAAACGTCTGCAT--------------------------- 1472

Qy      458 GluAsnAspMetTyrAlaGlyTyrHisGluSerAspThrSerGluAlaAlaAsnGluAla 477
Db     1473 ------------------GGCCTTGATGAAGAAGCAGAACAGAAACTTTTTCAGAGAAA 1514

Qy      478 LysIleGlnLysLeuLysGluLeuSerLysLeuProGlyIleTyrAspArgLeuSerArg 497
Db     1515 CGTGTGGAATTGCTTAAGGAACTTTCCAGGAAACCAGACATTTATGAGAGGCTTGCTTCA 1574

Qy      498 SerLeuAlaProSerIleTrpGluLeuGluAspIleLysLysGlyLeuLeuCysGlnLeu 517
Db     1575 GCCTTGGCTCCAAGCATTTATGAACATGAAGATATAAAGAAGGGAATTTTGCTTCAGCTC 1634

Qy      518 PheGlyGlyLysAlaLysLysIle------ProSerGlyAlaSerPheArgGlyAspIle 535
Db     1635 TTTGGCGGGACAAGGAAGGATTTTAGTCACACTGGAAGGGGCAAATTTCGGGCTGAGATC 1694

Qy      536 AsnValLeuLeuValGlyAspProGlyThrSerLysSerGlnLeuLeuGlnTyrValHis 555
Db     1695 AACATCTTGCTGTGTGGCGACCCTGGTACCAGCAAGTCCCAGCTGCTGCAGTACGTGTAC 1754

Qy      556 LysIleAlaProArgGlyIleTyrThrSerGlyArgGlySerSerAlaValGlyLeuThr 575
Db     1755 AACCTCGTCCCCAGGGGCCAGTACACGTCTGGGAAGGGCTCCAGTGCAGTTGGCCTCACT 1814

Qy      576 AlaTyrValThrLysAspProGluThrArgGluThrValLeuGluSerGlyAlaLeuVal 595
Db     1815 GCGTACGTAATGAAAGACCCTGAGACAAGGCAGCTGGTCCTGCAGACAGGTGCTCTTGTC 1874

Qy      596 LeuSerAspArgGlyIleCysCysIleAspGluPheAspLysMetSerAspAsnAlaArg 615
Db     1875 CTGAGTGACAACGGCATCTGCTGTATCGATGAGTTCGACAAGATGAATGAAAGTACAAGA 1934

Qy      616 SerMetLeuHisGluValMetGluGlnGlnThrValSerValAlaLysGlyGlyIleIle 635
Db     1935 TCGGTATTGCATGAAGTCATGGAACAGCAGACTCTGTCCATTGCAAAGGCTGGGATCATC 1994

Qy      636 AlaSerLeuAsnAlaArgThrSerValLeuAlaCysAlaAsnProSerGlySerArgTyr 655
Db     1995 TGTCAGCTCAATGCGCGCACCTCTGTCCTGGCAGCAGCAAATCCCATTGAGTCTCAGTGG 2054

Qy      656 AsnAlaArgLeuSerValIleAspAsnIleGlnLeuProProThrLeuLeuSerArgPhe 675
Db     2055 AATCCTAAAAAAACAACCATTGAAAACATCCAGCTGCCTCATACTTTATTATCAAGGTTT 2114

Qy      676 AspLeuIleTyrLeuMetLeuAspLysProAspGluGlnAsnAspArgArgLeuAlaArg 695
Db     2115 GATTTGATCTTCCTCATGCTGGACCCTCAGGACGAAGCCTATGACAGGCGTCTGGCTCAC 2174

Qy      696 HisLeuValAlaLeuHisTyrGluAsnTyrGluValSerLysGlnAspAlaLeuAspLeu 715
Db     2175 CACCTGGTCGCACTGTACTACCAGAGCGAGGAGCAGGCAGAGGAGGAGCTCCTGGACATG 2234

Qy      716 GlnThrLeuThrAlaTyrIleThrTyrAlaArgGlnHisValHisProThrLeuSerAsp 735
Db     2235 GCGGTGCTAAAGGACTACATTGCCTACGCGCACAGCACCATCATGCCGCGGCTAAGTGAG 2294

Qy      736 GluAlaAlaGluAspLeuIleAsnGlyTyrValGluMetArgGlnLysGlyAsnPhePro 755
```

```
Db     2295 GAAGCCAGCCAGGCTCTCATCGAGGCTTATGTAGACATGAGGAAG-----------ATT 2342

Qy      756 GlySerSerLysLysValIleThrAlaThrProArgGlnLeuGluSerMetIleArgIle 775
Db     2343 GGCAGTAGCCGGGGAATGGTTTCTGCATACCCTCGACAGCTAGAGTCATTAATCCGCTTA 2402

Qy      776 SerGluAlaLeuAlaArgMetArgPheSerGluValValGluLysValAspAlaAlaGlu 795
Db     2403 GCAGAAGCCCATGCTAAAGTAAGATTGTCTAACAAAGTTGAAGCCATTGATGTGGAAGAG 2462

Qy      796 AlaValArgLeuLeuAspValAlaLeuGlnGlnSerAlaThrAspHisAlaThrGlyThr 815
Db     2463 GCCAAACGCCTCCATCGGGAAGCTCTGAAGCAGTCTGCAACTGATCCCCGGACTGGCATC 2522

Qy      816 IleAspMetAspLeuIleThrThrGlyValSerAlaSerGluArgIleArgArgAlaAsn 835
Db     2523 GTGGACATATCTATTCTTACTACGGGGATGAGTGCCACCTCTCGTAAACGGAAAGAAGAA 2582

Qy      836 LeuLeuAlaAlaLeuArgGluLeuIleAlaAspLysIleSerProGlySerSerSerGly 855
Db     2583 TTAGCTGAAGCATTGAAAAAGCTTATT---------TTATCTAAGGGCAAAACACCAGCT 2633

Qy      856 LeuLysThrSerGlnLeuLeuGluAspIleArgSerGlnSerSerValAspValSerLeu 875
Db     2634 CTAAAATACCAGCAACTTTTGAAGATATTCGGGACAATCTGACATAGCAATTACTAAA 2693

Qy      876 GlnAspIleLysAsnAlaLeuGlySerLeuGlnGlyGluGlyPheLeuThrValHisGly 895
Db     2694 GATATGTTTGAAGAAGCACTGCGTGCCCTGGCAGATGATGATTTCCTGACAGTGACTGGG 2753

Qy      896 AspIleValLys 899
Db     2754 AAGACCGTGCGC 2765
```

```
RESULT 14
US-09-954-456-569
; Sequence 569, Application US/09954456
; Patent No. US20020115057A1
; GENERAL INFORMATION:
;  APPLICANT: Young, Paul
;  TITLE OF INVENTION: Process for Identifying Anti-Cancer Therapeutic Agents Using Canc
;  TITLE OF INVENTION:  Sets
;  FILE REFERENCE: 689290-76
;  CURRENT APPLICATION NUMBER: US/09/954,456
;  CURRENT FILING DATE:  2001-09-18
;  PRIOR APPLICATION NUMBER: US/60/233,617
;  PRIOR FILING DATE: 2000-09-18
;  PRIOR APPLICATION NUMBER: US/60/234,052
;  PRIOR FILING DATE: 2000-09-20
;  PRIOR APPLICATION NUMBER: US/60/234,923
;  PRIOR FILING DATE: 2000-09-25
;  PRIOR APPLICATION NUMBER: US/60/235,134
;  PRIOR FILING DATE: 2000-09-25
;  PRIOR APPLICATION NUMBER: US/60/235,637
;  PRIOR FILING DATE: 2000-09-26
;  PRIOR APPLICATION NUMBER: US/60/235,638
;  PRIOR FILING DATE: 2000-09-26
;  PRIOR APPLICATION NUMBER: US/60/235,711
;  PRIOR FILING DATE: 2000-09-27
;  PRIOR APPLICATION NUMBER: US/60/235,720
;  PRIOR FILING DATE: 2000-09-27
;  PRIOR APPLICATION NUMBER: US/60/235,840
;  PRIOR FILING DATE: 2000-09-27
;  PRIOR APPLICATION NUMBER: US/60/235,863
;  PRIOR FILING DATE: 2000-09-27
;  NUMBER OF SEQ ID NOS: 2276
;  SOFTWARE: PatentIn version 3.0
; SEQ ID NO 569
;   LENGTH: 3273
;   TYPE: DNA
;   ORGANISM: Homo sapiens
US-09-954-456-569
```

```
Alignment Scores:
Pred. No.:              1.05e-145     Length:        3273
Score:                  1817.50       Matches:       406
Percent Similarity:     57.78%        Conservative:  151
Best Local Similarity:  42.12%        Mismatches:    256
Query Match:            39.43%        Indels:        151
DB:                     3             Gaps:          18

US-10-768-511-8 (1-901) x US-09-954-456-569 (1-3273)

Qy       38 AlaAlaSerProValAlaGlyArgArgAlaValArgGlnThrProThrSerAlaValArg 57
Db       21 GCGACGTCGTCGGCAAGCGGCCGC---------------CTTCCACGTAACGCG---CGC 62

Qy       58 ArgArgGlyArgGluThrAspSerAlaArgArgArgArgSerArg--------------- 72
Db       63 CGGCGGGGGAGGGCGTTGGCGCGGAGCCGACGGGAACGTCCGCGCTGCGGAGCAGGGCAG 122

Qy       73 ---SerArgSerLeuGlyAsnSerValTyrSerSerProTyrAspAlaGlyThrProGly 91
Db      123 GGAAGCCGGGAGGCGGGC---CCGGCCCGAGCTTGTCCTTGTCGCGCAGGTACTCCGAGC 179

Qy       92 ThrProGlyThrProValAlaThrProVal--------------------TyrAlaThr 104
Db      180 ACTATGTCGTCCCCGGCGTCGACCCCGAGCCGCCGCGGCAGCCGGCGTGGAAGGGCCACC 239

Qy      105 ProValGlyThrPro--------------------------------------------- 109
Db      240 CCCGCCCAGACGCCTCGGAGTGAGGATGCCAGGTCATCTCCCTCTCAGAGACGTAGAGGC 299

Qy      109 ------------------------------------------------------------ 109
Db      300 GAGGATTCCACCTCCACGGGGAGTTGCAGCCGATGCCAACCTCGCCTGGAGTGGACCTG 359

Qy      109 ------------------------------------------------------------ 109
Db      360 CAGAGCACTGCTGCGCAGGACGTGCTGTTTTCCAGCCCTCCCCAAATGCATTCTTCAGCT 419

Qy      110 ------------------------------------MetGlyThrProSerPheHis--- 116
Db      420 ATCCCTCTTGACTTTGATGTTAGTTCACCACTGACATACGGCACTCCCAGCTCTCGGGTA 479

Qy      117 ------------------------ArgGlyThrProGlnTyrLysGlnArgSerGluLeu 128
Db      480 GAGGGAACCCCAAGAAGTGGTGTTAGGGGCACACCT---GTGAGACAGAGGCCTGACCTG 536

Qy      129 GlySerGlnGlyLysProLeuHisArgArgArgArgSerGlnSerArgGluProGlyHis 148
Db      537 GGCTCTGCACAGAAGGGCCTGCAAGTGGATCTGCAGTCT-------------------- 575

Qy      149 ArgSerProSerArgGluProSerAlaAspGlyArgProSerGluSerAlaGluProAsp 168
Db      576 --------------------------GACGGGGCAGCAGCAGAAGATATAGTGGCAAGT 608

Qy      169 AspThrLeuGlyGlyGluTyrAlaTyrValTrpGlyThrAsnValAsnIleProAspVal 188
Db      609 GAGCAGTCTCTAGGCCAAAAACTTGTGATCTGGGGAACAGATGTAAATGTGGCAGCATGC 668

Qy      189 LeuArgAlaIleArgArgPheLeuHisAsnTyr----------------------Arg 200
Db      669 AAAGAAAACTTTCAGAGATTTCTTCAGCGTTTTATTGACCCTCTGGCTAAAGAAGAAGAA 728

Qy      201 SerSerAlaHisAspLeuAsnSerLys---TyrIleGlnIleIleGluGluThrValGlu 219
Db      729 AATGTTGGCATAGATATTACTGAACCTCTATACATGCAACGACTTGGGGAGATTAATGTT 788

Qy      220 ArgGluGluAspThrLeuAsnIleAspMetSerAspIleTyrAspHisAspProAspLeu 239
Db      789 ATTGGTGAGCAATTTTTAAATGTGAACTGTGAACACATCAAATCATTTGACAAAAATTTG 848

Qy      240 TyrAlaLysIleValArgTyrProLeuAspIleIleProLeuLeuAspThrGluCysGln 259
Db      849 TACAGACAACTCATCTCTTACCCACAGGAAGTTATTCCAACTTTTGACATGGCTGTCAAT 908
```

```
Qy        260 GluValAlaThrSerLeuLeuPro------ThrPheGluLysHisIleGluAlaArgPro 277
Db        909 GAAATCTTCTTTGACCGTTACCCTGACTCAATCTTAGAACATCAGATTCAAGTAAGACCA 968

Qy        278 PheAsnLeuLysAlaSerValHisMetArgGluLeuAsnProSerAspIleAspLysLeu 297
Db        969 TTCAACGCATTGAAGACTAAGAATATGAGAAACCTGAATCCAGAAGACATTGACCAGCTC 1028

Qy        298 ValSerValLysGlyMetValIleArgCysSerSerIleIleProGluIleLysGlyAla 317
Db       1029 ATCACCATCAGCGGCATGGTGATCAGGACATCCCAGCTGATTCCCGAGATGCAGGAGGCC 1088

Qy        318 PhePheLysCysLeuValCysGlyHisSerProProLeuValThrValValLysGlyArg 337
Db       1089 TTCTTCCAGTGCCAAGTGTGTGCCCACACG---ACCCGGGTGGAGATGGACCGCGGCCGC 1145

Qy        338 ValGluGluProThrArgCysGluLysProGluCysAlaAlaArgAsnAlaMetSerLeu 357
Db       1146 ATTGCAGAGCCCAGTGTGTGCGGGCGC------TGCCACACCACCCACAGCATGGCACTC 1199

Qy        358 IleHisAsnArgCysThrPheAlaAsnLysGlnIleValArgLeuGlnGluThrProAsp 377
Db       1200 ATCCACAACCGCTCCCTCTTCTCTGACAAGCAGATGATCAAGCTTCAGGAGTCTCCGGAA 1259

Qy        378 AlaIleProGluGlyGluThrProHisThrValSerMetCysLeuTyrAsnThrMetVal 397
Db       1260 GACATGCCTGCAGGGCAGACACCACACACAGTTATCCTGTTTGCTCACAATGATCTCGTT 1319

Qy        398 AspAlaValLysProGlyAspArgIleGluValThrGlyValPheLysAlaMetAlaVal 417
Db       1320 GACAAGGTCCAGCCTGGGGACAGAGTGAATGTTACAGGCATCTATCGAGCTGTGCCTATT 1379

Qy        418 ArgValGlyProAsnGlnArgThrLeuArgAlaLeuTyrLysThrTyrIleAspCysVal 437
Db       1380 CGAGTCAATCCAAGAGTGAGTAATGTGAAGTCTGTCTACAAAACCCACATTGATGTCATT 1439

Qy        438 HisValLysLysSerAspArgGlyArgLeuGlnThrGluAspProMetGluMetAspLys 457
Db       1440 CATTATCGGAAAACGGATGCAAAACGTCTGCAT--------------------------- 1472

Qy        458 GluAsnAspMetTyrAlaGlyTyrHisGluSerAspThrSerGluAlaAlaAsnGluAla 477
Db       1473 -----------------GGCCTTGATGAAGAAGCAGAACAGAACTTTTTCAGAGAAA 1514

Qy        478 LysIleGlnLysLeuLysGluLeuSerLysLeuProGlyIleTyrAspArgLeuSerArg 497
Db       1515 CGTGTGGAATTGCTTAAGGAACTTTCCAGGAAACCAGACATTTATGAGAGGCTTGCTTCA 1574

Qy        498 SerLeuAlaProSerIleTrpGluLeuGluAspIleLysLysGlyLeuLeuCysGlnLeu 517
Db       1575 GCCTTGGCTCCAAGCATTTATGAACATGAAGATATAAAGAAGGGAATTTTGCTTCAGCTC 1634

Qy        518 PheGlyGlyLysAlaLysLysIle------ProSerGlyAlaSerPheArgGlyAspIle 535
Db       1635 TTTGGCGGGACAAGGAAGGATTTTAGTCACACTGGAAGGGGCAAATTTCGGGCTGAGATC 1694

Qy        536 AsnValLeuLeuValGlyAspProGlyThrSerLysSerGlnLeuLeuGlnTyrValHis 555
Db       1695 AACATCTTGCTGTGTGGCGACCCTGGTACCAGCAAGTCCCAGCTGCTGCAGTACGTGTAC 1754

Qy        556 LysIleAlaProArgGlyIleTyrThrSerGlyArgGlySerSerAlaValGlyLeuThr 575
Db       1755 AACCTCGTCCCCAGGGGCCAGTACACGTCTGGGAAGGGCTCCAGTGCAGTTGGCCTCACT 1814

Qy        576 AlaTyrValThrLysAspProGluThrArgGluThrValLeuGluSerGlyAlaLeuVal 595
Db       1815 GCGTACGTAATGAAAGACCCTGAGACAAGGCAGCTGGTCCTGCAGACAGGTGCTCTTGTC 1874

Qy        596 LeuSerAspArgGlyIleCysCysIleAspGluPheAspLysMetSerAspAsnAlaArg 615
Db       1875 CTGAGTGACAACGGCATCTGCTGTATCGATGAGTTCGACAAGATGAATGAAAGTACAAGA 1934

Qy        616 SerMetLeuHisGluValMetGluGlnGlnThrValSerValAlaLysGlyGlyIleIle 635
Db       1935 TCGGTATTGCATGAAGTCATGGAACAGCAGACTCTGTCCATTGCAAAGGCTGGGATCATC 1994

Qy        636 AlaSerLeuAsnAlaArgThrSerValLeuAlaCysAlaAsnProSerGlySerArgTyr 655
Db       1995 TGTCAGCTCAATGCGCGCACCTCTGTCCTGGCAGCAGCAAATCCCATTGAGTCTCAGTGG 2054

Qy        656 AsnAlaArgLeuSerValIleAspAsnIleGlnLeuProProThrLeuLeuSerArgPhe 675
Db       2055 AATCCTAAAAAAACAACCATTGAAAACATCCAGCTGCCTCATACTTTATTATCAAGGTTT 2114

Qy        676 AspLeuIleTyrLeuMetLeuAspLysProAspGluGlnAsnAspArgArgLeuAlaArg 695
Db       2115 GATTTGATCTTCCTCATGCTGGACCCTCAGGACGAAGCCTATGACAGGCGTCTGGCTCAC 2174

Qy        696 HisLeuValAlaLeuHisTyrGluAsnTyrGluValSerLysGlnAspAlaLeuAspLeu 715
Db       2175 CACCTGGTCGCACTGTACTACCAGAGCGAGGAGCAGGCAGAGGAGGAGCTCCTGGACATG 2234

Qy        716 GlnThrLeuThrAlaTyrIleThrTyrAlaArgGlnHisValHisProThrLeuSerAsp 735
Db       2235 GCGGTGCTAAAGGACTACATTGCCTACGCGCACAGCACCATCATGCCGCGGCTAAGTGAG 2294

Qy        736 GluAlaAlaGluAspLeuIleAsnGlyTyrValGluMetArgGlnLysGlyAsnPhePro 755
Db       2295 GAAGCCAGCCAGGCTCTCATCGAGGCTTATGTAGACATGAGGAAG------------ATT 2342

Qy        756 GlySerSerLysLysValIleThrAlaThrProArgGlnLeuGluSerMetIleArgIle 775
Db       2343 GGCAGTAGCCGGGGAATGGTTTCTGCATACCCTCGACAGCTAGAGTCATTAATCCGCTTA 2402

Qy        776 SerGluAlaLeuAlaArgMetArgPheSerGluValValGluLysValAspAlaAlaGlu 795
Db       2403 GCAGAAGCCCATGCTAAAGTAAGATTGTCTAACAAAGTTGAAGCCATTGATGTGGAAGAG 2462

Qy        796 AlaValArgLeuLeuAspValAlaLeuGlnGlnSerAlaThrAspHisAlaThrGlyThr 815
Db       2463 GCCAAACGCCTCCATCGGGAAGCTCTGAAGCAGTCTGCAACTGATCCCCGGACTGGCATC 2522

Qy        816 IleAspMetAspLeuIleThrThrGlyValSerAlaSerGluArgIleArgArgAlaAsn 835
Db       2523 GTGGACATATCTATTCTTACTACGGGGATGAGTGCCACCTCTCGTAAACGGAAAGAAGAA 2582

Qy        836 LeuLeuAlaAlaLeuArgGluLeuIleAlaAspLysIleSerProGlySerSerSerGly 855
Db       2583 TTAGCTGAAGCATTGAAAAAGCTTATT---------TTATCTAAGGGCAAAACACCAGCT 2633

Qy        856 LeuLysThrSerGlnLeuLeuGluAspIleArgSerGlnSerSerValAspValSerLeu 875
Db       2634 CTAAAATACCAGCAACTTTTTGAAGATATTCGGGGACAATCTGACATAGCAATTACTAAA 2693

Qy        876 GlnAspIleLysAsnAlaLeuGlySerLeuGlnGlyGluGlyPheLeuThrValHisGly 895
Db       2694 GATATGTTTGAAGAAGCACTGCGTGCCCTGGCAGATGATGATTTCCTGACAGTGACTGGG 2753

Qy        896 AspIleValLys 899
Db       2754 AAGACCGTGCGC 2765
```

RESULT 15
US-09-954-456-976
; Sequence 976, Application US/09954456
; Patent No. US20020115057A1
; GENERAL INFORMATION:
; APPLICANT: Young, Paul
; TITLE OF INVENTION: Process for Identifying Anti-Cancer Therapeutic Agents Using Canc
; TITLE OF INVENTION: Sets
; FILE REFERENCE: 689290-76
; CURRENT APPLICATION NUMBER: US/09/954,456
; CURRENT FILING DATE: 2001-09-18
; PRIOR APPLICATION NUMBER: US/60/233,617
; PRIOR FILING DATE: 2000-09-18

```
;  PRIOR APPLICATION NUMBER: US/60/234,052
;  PRIOR FILING DATE: 2000-09-20
;  PRIOR APPLICATION NUMBER: US/60/234,923
;  PRIOR FILING DATE: 2000-09-25
;  PRIOR APPLICATION NUMBER: US/60/235,134
;  PRIOR FILING DATE: 2000-09-25
;  PRIOR APPLICATION NUMBER: US/60/235,637
;  PRIOR FILING DATE: 2000-09-26
;  PRIOR APPLICATION NUMBER: US/60/235,638
;  PRIOR FILING DATE: 2000-09-26
;  PRIOR APPLICATION NUMBER: US/60/235,711
;  PRIOR FILING DATE: 2000-09-27
;  PRIOR APPLICATION NUMBER: US/60/235,720
;  PRIOR FILING DATE: 2000-09-27
;  PRIOR APPLICATION NUMBER: US/60/235,840
;  PRIOR FILING DATE: 2000-09-27
;  PRIOR APPLICATION NUMBER: US/60/235,863
;  PRIOR FILING DATE: 2000-09-27
;  NUMBER OF SEQ ID NOS: 2276
;  SOFTWARE: PatentIn version 3.0
; SEQ ID NO 976
;   LENGTH: 3273
;   TYPE: DNA
;   ORGANISM: Homo sapiens
US-09-954-456-976

Alignment Scores:
Pred. No.:              1.05e-145     Length:        3273
Score:                  1817.50       Matches:       406
Percent Similarity:     57.78%        Conservative:  151
Best Local Similarity:  42.12%        Mismatches:    256
Query Match:            39.43%        Indels:        151
DB:                     3             Gaps:          18

US-10-768-511-8 (1-901) x US-09-954-456-976 (1-3273)

Qy     38 AlaAlaSerProValAlaGlyArgArgAlaValArgGlnThrProThrSerAlaValArg 57
Db     21 GCGACGTCGTCGGCAAGCGGCCGC---------------CTTCCACGTAACGCG---CGC 62

Qy     58 ArgArgGlyArgGluThrAspSerAlaArgArgArgArgSerArg--------------- 72
Db     63 CGGCGGGGGAGGGCGTTGGCGCGGAGCCGACGGGAACGTCCGCGCTGCGGAGCAGGGCAG 122

Qy     73 ---SerArgSerLeuGlyAsnSerValTyrSerSerProTyrAspAlaGlyThrProGly 91
Db    123 GGAAGCCGGGAGGCGGGC---CCGGCCCGAGCTTGTCCTTGTCGCGCAGGTACTCCGAGC 179

Qy     92 ThrProGlyThrProValAlaThrProVal--------------------TyrAlaThr 104
Db    180 ACTATGTCGTCCCCGGCGTCGACCCCGAGCCGCCGCGGCAGCCGGCGTGGAAGGGCCACC 239

Qy    105 ProValGlyThrPro--------------------------------------------- 109
Db    240 CCCGCCCAGACGCCTCGGAGTGAGGATGCCAGGTCATCTCCCTCTCAGAGACGTAGAGGC 299

Qy    109 ------------------------------------------------------------ 109
Db    300 GAGGATTCCACCTCCACGGGGGAGTTGCAGCCGATGCCAACCTCGCCTGGAGTGGACCTG 359

Qy    109 ------------------------------------------------------------ 109
Db    360 CAGAGCACTGCTGCGCAGGACGTGCTGTTTTCCAGCCCTCCCCAAATGCATTCTTCAGCT 419

Qy    110 ------------------------------------MetGlyThrProSerPheHis--- 116
Db    420 ATCCCTCTTGACTTTGATGTTAGTTCACCACTGACATACGGCACTCCCAGCTCTCGGGTA 479

Qy    117 ------------------------ArgGlyThrProGlnTyrLysGlnArgSerGluLeu 128
Db    480 GAGGGAACCCCAAGAAGTGGTGTTAGGGGCACACCT---GTGAGACAGAGGCCTGACCTG 536

Qy    129 GlySerGlnGlyLysProLeuHisArgArgArgArgSerGlnSerArgGluProGlyHis 148
Db    537 GGCTCTGCACAGAAGGGCCTGCAAGTGGATCTGCAGTCT-------------------- 575

Qy    149 ArgSerProSerArgGluProSerAlaAspGlyArgProSerGluSerAlaGluProAsp 168
Db    576 --------------------------GACGGGGCAGCAGCAGAAGATATAGTGGCAAGT 608

Qy    169 AspThrLeuGlyGlyGluTyrAlaTyrValTrpGlyThrAsnValAsnIleProAspVal 188
Db    609 GAGCAGTCTCTAGGCCAAAAACTTGTGATCTGGGGAACAGATGTAAATGTGGCAGCATGC 668

Qy    189 LeuArgAlaIleArgArgPheLeuHisAsnTyr-----------------------Arg 200
Db    669 AAAGAAAACTTTCAGAGATTTCTTCAGCGTTTTATTGACCCTCTGGCTAAAGAAGAAGAA 728

Qy    201 SerSerAlaHisAspLeuAsnSerLys---TyrIleGlnIleIleGluGluThrValGlu 219
Db    729 AATGTTGGCATAGATATTACTGAACCTCTATACATGCAACGACTTGGGGAGATTAATGTT 788

Qy    220 ArgGluGluAspThrLeuAsnIleAspMetSerAspIleTyrAspHisAspProAspLeu 239
Db    789 ATTGGTGAGCAATTTTTAAATGTGAACTGTGAACACATCAAATCATTTGACAAAAATTTG 848

Qy    240 TyrAlaLysIleValArgTyrProLeuAspIleIleProLeuLeuAspThrGluCysGln 259
Db    849 TACAGACAACTCATCTCTTACCCACAGGAAGTTATTCCAACTTTTGACATGGCTGTCAAT 908

Qy    260 GluValAlaThrSerLeuLeuPro------ThrPheGluLysHisIleGluAlaArgPro 277
Db    909 GAAATCTTCTTTGACCGTTACCCTGACTCAATCTTAGAACATCAGATTCAAGTAAGACCA 968

Qy    278 PheAsnLeuLysAlaSerValHisMetArgGluLeuAsnProSerAspIleAspLysLeu 297
Db    969 TTCAACGCATTGAAGACTAAGAATATGAGAAACCTGAATCCAGAAGACATTGACCAGCTC 1028

Qy    298 ValSerValLysGlyMetValIleArgCysSerSerIleIleProGluIleLysGlyAla 317
Db   1029 ATCACCATCAGCGGCATGGTGATCAGGACATCCCAGCTGATTCCCGAGATGCAGGAGGCC 1088

Qy    318 PhePheLysCysLeuValCysGlyHisSerProProLeuValThrValValLysGlyArg 337
Db   1089 TTCTTCCAGTGCCAAGTGTGTGCCCACACG---ACCCGGGTGGAGATGGACCGCGGCCGC 1145

Qy    338 ValGluGluProThrArgCysGluLysProGluCysAlaAlaArgAsnAlaMetSerLeu 357
Db   1146 ATTGCAGAGCCCAGTGTGTGCGGGCGC------TGCCACACCACCCACAGCATGGCACTC 1199

Qy    358 IleHisAsnArgCysThrPheAlaAsnLysGlnIleValArgLeuGlnGluThrProAsp 377
Db   1200 ATCCACAACCGCTCCCTCTTCTCTGACAAGCAGATGATCAAGCTTCAGGAGTCTCCGGAA 1259

Qy    378 AlaIleProGluGlyGluThrProHisThrValSerMetCysLeuTyrAsnThrMetVal 397
Db   1260 GACATGCCTGCAGGGCAGACACCACACACAGTTATCCTGTTTGCTCACAATGATCTCGTT 1319

Qy    398 AspAlaValLysProGlyAspArgIleGluValThrGlyValPheLysAlaMetAlaVal 417
Db   1320 GACAAGGTCCAGCCTGGGGACAGAGTGAATGTTACAGGCATCTATCGAGCTGTGCCTATT 1379

Qy    418 ArgValGlyProAsnGlnArgThrLeuArgAlaLeuTyrLysThrTyrIleAspCysVal 437
Db   1380 CGAGTCAATCCAAGAGTGAGTAATGTGAAGTCTGTCTACAAAACCCACATTGATGTCATT 1439

Qy    438 HisValLysLysSerAspArgGlyArgLeuGlnThrGluAspProMetGluMetAspLys 457
Db   1440 CATTATCGGAAAACGGATGCAAAACGTCTGCAT------------------------- 1472

Qy    458 GluAsnAspMetTyrAlaGlyTyrHisGluSerAspThrSerGluAlaAlaAsnGluAla 477
Db   1473 ------------------GGCCTTGATGAAGAAGCAGAACAGAAACTTTTTCAGAGAAA 1514

Qy    478 LysIleGlnLysLeuLysGluLeuSerLysLeuProGlyIleTyrAspArgLeuSerArg 497
```

```
Db      1515 CGTGTGGAATTGCTTAAGGAACTTTCCAGGAAACCAGACATTTATGAGAGGCTTGCTTCA 1574

Qy       498 SerLeuAlaProSerIleTrpGluLeuGluAspIleLysLysGlyLeuLeuCysGlnLeu 517
Db      1575 GCCTTGGCTCCAAGCATTTATGAACATGAAGATATAAAGAAGGGAATTTTGCTTCAGCTC 1634

Qy       518 PheGlyGlyLysAlaLysLysIle------ProSerGlyAlaSerPheArgGlyAspIle 535
Db      1635 TTTGGCGGGACAAGGAAGGATTTTAGTCACACTGGAAGGGGCAAATTTCGGGCTGAGATC 1694

Qy       536 AsnValLeuLeuValGlyAspProGlyThrSerLysSerGlnLeuLeuGlnTyrValHis 555
Db      1695 AACATCTTGCTGTGTGGCGACCCTGGTACCAGCAAGTCCCAGCTGCTGCAGTACGTGTAC 1754

Qy       556 LysIleAlaProArgGlyIleTyrThrSerGlyArgGlySerSerAlaValGlyLeuThr 575
Db      1755 AACCTCGTCCCCAGGGGCCAGTACACGTCTGGGAAGGGCTCCAGTGCAGTTGGCCTCACT 1814

Qy       576 AlaTyrValThrLysAspProGluThrArgGluThrValLeuGluSerGlyAlaLeuVal 595
Db      1815 GCGTACGTAATGAAAGACCCTGAGACAAGGCAGCTGGTCCTGCAGACAGGTGCTCTTGTC 1874

Qy       596 LeuSerAspArgGlyIleCysCysIleAspGluPheAspLysMetSerAspAsnAlaArg 615
Db      1875 CTGAGTGACAACGGCATCTGCTGTATCGATGAGTTCGACAAGATGAATGAAAGTACAAGA 1934

Qy       616 SerMetLeuHisGluValMetGluGlnGlnThrValSerValAlaLysGlyGlyIleIle 635
Db      1935 TCGGTATTGCATGAAGTCATGGAACAGCAGACTCTGTCCATTGCAAAGGCTGGGATCATC 1994

Qy       636 AlaSerLeuAsnAlaArgThrSerValLeuAlaCysAlaAsnProSerGlySerArgTyr 655
Db      1995 TGTCAGCTCAATGCGCGCACCTCTGTCCTGGCAGCAGCAAATCCCATTGAGTCTCAGTGG 2054

Qy       656 AsnAlaArgLeuSerValIleAspAsnIleGlnLeuProProThrLeuLeuSerArgPhe 675
Db      2055 AATCCTAAAAAAACAACCATTGAAAACATCCAGCTGCCTCATACTTTATTATCAAGGTTT 2114

Qy       676 AspLeuIleTyrLeuMetLeuAspLysProAspGluGlnAsnAspArgArgLeuAlaArg 695
Db      2115 GATTTGATCTTCCTCATGCTGGACCCTCAGGACGAAGCCTATGACAGGCGTCTGGCTCAC 2174

Qy       696 HisLeuValAlaLeuHisTyrGluAsnTyrGluValSerLysGlnAspAlaLeuAspLeu 715
Db      2175 CACCTGGTCGCACTGTACTACCAGAGCGAGGAGCAGGCAGAGGAGGAGCTCCTGGACATG 2234

Qy       716 GlnThrLeuThrAlaTyrIleThrTyrAlaArgGlnHisValHisProThrLeuSerAsp 735
Db      2235 GCGGTGCTAAAGGACTACATTGCCTACGCGCACAGCACCATCATGCCGCGGCTAAGTGAG 2294

Qy       736 GluAlaAlaGluAspLeuIleAsnGlyTyrValGluMetArgGlnLysGlyAsnPhePro 755
Db      2295 GAAGCCAGCCAGGCTCTCATCGAGGCTTATGTAGACATGAGGAAG------------ATT 2342

Qy       756 GlySerSerLysLysValIleThrAlaThrProArgGlnLeuGluSerMetIleArgIle 775
Db      2343 GGCAGTAGCCGGGGAATGGTTTCTGCATACCCTCGACAGCTAGAGTCATTAATCCGCTTA 2402

Qy       776 SerGluAlaLeuAlaArgMetArgPheSerGluValValGluLysValAspAlaAlaGlu 795
Db      2403 GCAGAAGCCCATGCTAAAGTAAGATTGTCTAACAAAGTTGAAGCCATTGATGTGGAAGAG 2462

Qy       796 AlaValArgLeuLeuAspValAlaLeuGlnGlnSerAlaThrAspHisAlaThrGlyThr 815
Db      2463 GCCAAACGCCTCCATCGGGAAGCTCTGAAGCAGTCTGCAACTGATCCCCGGACTGGCATC 2522

Qy       816 IleAspMetAspLeuIleThrThrGlyValSerAlaSerGluArgIleArgArgAlaAsn 835
Db      2523 GTGGACATATCTATTCTTACTACGGGGATGAGTGCCACCTCTCGTAAACGGAAAGAAGAA 2582

Qy       836 LeuLeuAlaAlaLeuArgGluLeuIleAlaAspLysIleSerProGlySerSerSerGly 855
Db      2583 TTAGCTGAAGCATTGAAAAAGCTTATT---------TTATCTAAGGGCAAAACACCAGCT 2633

Qy       856 LeuLysThrSerGlnLeuLeuGluAspIleArgSerGlnSerSerValAspValSerLeu 875
Db      2634 CTAAAATACCAGCAACTTTTGAAGATATTCGGGGACAATCTGACATAGCAATTACTAAA 2693

Qy       876 GlnAspIleLysAsnAlaLeuGlySerLeuGlnGlyGluGlyPheLeuThrValHisGly 895
Db      2694 GATATGTTTGAAGAAGCACTGCGTGCCCTGGCAGATGATGATTTCCTGACAGTGACTGGG 2753

Qy       896 AspIleValLys 899
Db      2754 AAGACCGTGCGC 2765
```

Search completed: December 6, 2005, 14:22:51
Job time : 1247 secs

```
OM protein - nucleic search, using frame_plus_p2n model

Run on:         December  6, 2005, 10:15:34 ; Search time 155 Seconds
                                              (without alignments)
                                              1809.092 Million cell updates/sec

Title:          US-10-768-511-8
Perfect score:  4609
Sequence:       1 MENNDALDIGAVSSPYPSQS..........LGSLQGEGFLTVHGDIVKRV 901

Scoring table:  BLOSUM62
                Xgapop 10.0 , Xgapext  0.5
                Ygapop 10.0 , Ygapext  0.5
                Fgapop  6.0 , Fgapext  7.0
                Delop   6.0 , Delext   7.0

Searched:       3289935 seqs, 155610033 residues

Total number of hits satisfying chosen parameters:      6579870

Minimum DB seq length: 0
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%
                 Maximum Match 100%
                 Listing first 45 summaries

Command line parameters:
-MODEL=frame+_p2n.model -DEV=xlh
-Q=/cgn2_1/USPTO_spool/US10768511/runat_02122005_102336_18498/app_query.fasta_1.1095
-DB=Published_Applications_NA_New -QFMT=fastap -SUFFIX=rnpbn -MINMATCH=0.1
-LOOPCL=0 -LOOPEXT=0 -UNITS=bits -START=1 -END=-1 -MATRIX=blosum62
-TRANS=human40.cdi -LIST=45 -DOCALIGN=200 -THR_SCORE=pct -THR_MAX=100
-THR_MIN=0 -ALIGN=15 -MODE=LOCAL -OUTFMT=pto -NORM=ext -HEAPSIZE=500 -MINLEN=0
-MAXLEN=2000000000 -USER=US10768511_@CGN_1_1_117_@runat_02122005_102336_18498
-NCPU=6 -ICPU=3 -NO_MMAP -LARGEQUERY -NEG_SCORES=0 -WAIT -DSPBLOCK=100
-LONGLOG -DEV_TIMEOUT=120 -WARN_TIMEOUT=30 -THREADS=1 -XGAPOP=10 -XGAPEXT=0.5
-FGAPOP=6 -FGAPEXT=7 -YGAPOP=10 -YGAPEXT=0.5 -DELOP=6 -DELEXT=7

Database :      Published_Applications_NA_New:*
                1:  /cgn2_6/ptodata/1/pubpna/US09_NEW_PUB.seq:*
                2:  /cgn2_6/ptodata/1/pubpna/US06_NEW_PUB.seq:*
                3:  /cgn2_6/ptodata/1/pubpna/US07_NEW_PUB.seq:*
                4:  /cgn2_6/ptodata/1/pubpna/US08_NEW_PUB.seq:*
                5:  /cgn2_6/ptodata/1/pubpna/PCT_NEW_PUB.seq:*
                6:  /cgn2_6/ptodata/1/pubpna/US10_NEW_PUB.seq:*
                7:  /cgn2_6/ptodata/1/pubpna/US11_NEW_PUB.seq:*
                8:  /cgn2_6/ptodata/1/pubpna/US11_NEW_PUB.seq2:*
                9:  /cgn2_6/ptodata/1/pubpna/US11_NEW_PUB.seq3:*
                10: /cgn2_6/ptodata/1/pubpna/US60_NEW_PUB.seq:*
```

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

	Result No.	Score	% Query Match	Length	DB	ID	Description
	1	1016	22.0	2821	7	US-11-087-227-13	Sequence 13, Appl
	2	974	21.1	2900	7	US-11-087-227-15	Sequence 15, Appl
	3	943.5	20.5	3453	7	US-11-087-227-11	Sequence 11, Appl
	4	930	20.2	2715	6	US-10-507-275-4	Sequence 4, Appli
	5	918.5	19.9	3769	7	US-11-087-227-89	Sequence 89, Appl
	6	856	18.6	3379	6	US-10-507-275-2	Sequence 2, Appli
	7	438.5	9.5	507	7	US-11-108-172-158	Sequence 158, App
	8	176.5	3.8	1312	6	US-10-750-185-55669	Sequence 55669, A
	9	174	3.8	1851	6	US-10-750-185-45476	Sequence 45476, A
	10	150	3.3	5253	9	US-11-004-057-3	Sequence 3, Appli
	11	144	3.1	876	6	US-10-750-185-27656	Sequence 27656, A
	12	141.5	3.1	2211	6	US-10-467-657-3451	Sequence 3451, Ap
c	13	137	3.0	33014	7	US-11-077-716-1	Sequence 1, Appli
c	14	137	3.0	34555	6	US-10-623-155-479	Sequence 479, App
	15	129	2.8	1260	7	US-11-074-176-47	Sequence 47, Appl
	16	129	2.8	1494	6	US-10-467-657-7809	Sequence 7809, Ap
	17	129	2.8	2736	6	US-10-858-730-38	Sequence 38, Appl
	18	129	2.8	3513	6	US-10-467-657-7841	Sequence 7841, Ap
	19	128.5	2.8	6012	6	US-10-467-657-83	Sequence 83, Appl
	20	128.5	2.8	6012	6	US-10-467-657-6321	Sequence 6321, Ap
c	21	127.5	2.8	5253	6	US-10-821-234-293	Sequence 293, App
	22	122.5	2.7	1723	7	US-11-051-568-28	Sequence 28, Appl
	23	122.5	2.7	6160	7	US-11-060-005-3	Sequence 3, Appli
	24	122	2.6	2828	6	US-10-957-569-30	Sequence 30, Appl
c	25	121.5	2.6	1590	6	US-10-858-730-162	Sequence 162, App
	26	121	2.6	3408	6	US-10-858-730-40	Sequence 40, Appl
	27	120.5	2.6	6210	7	US-11-056-470-1	Sequence 1, Appli
	28	120.5	2.6	15071	6	US-10-793-626-2963	Sequence 2963, Ap
	29	120	2.6	1593	6	US-10-858-730-138	Sequence 138, App
	30	119	2.6	2947	6	US-10-793-626-3805	Sequence 3805, Ap
	31	119	2.6	3513	6	US-10-858-730-142	Sequence 142, App
	32	119	2.6	3730	6	US-10-793-626-3499	Sequence 3499, Ap
	33	118	2.6	1893	6	US-10-131-826A-89	Sequence 89, Appl
	34	118	2.6	1893	7	US-11-102-240-149	Sequence 149, App
	35	118	2.6	3289	6	US-10-971-560-8	Sequence 8, Appli
c	36	118	2.6	3569	6	US-10-793-626-4304	Sequence 4304, Ap
	37	118	2.6	3911	9	US-11-004-057-5	Sequence 5, Appli
	38	118	2.6	4158	6	US-10-971-560-3	Sequence 3, Appli
	39	118	2.6	6458	6	US-10-971-560-1	Sequence 1, Appli
	40	118	2.6	9471	6	US-10-821-234-410	Sequence 410, App
c	41	117	2.5	3146	6	US-10-793-626-3375	Sequence 3375, Ap
	42	116.5	2.5	1152	6	US-10-467-657-507	Sequence 507, App
	43	116.5	2.5	2229	6	US-10-467-657-4081	Sequence 4081, Ap
	44	116	2.5	14770	6	US-10-821-234-268	Sequence 268, App
	45	115	2.5	2475	7	US-11-074-176-19	Sequence 19, Appl

ALIGNMENTS

```
RESULT 1
US-11-087-227-13
; Sequence 13, Application US/11087227
; Publication No. US20050260566A1
; GENERAL INFORMATION:
;  APPLICANT: Fischer, Timothy J.
;  APPLICANT:  Malinowski, Douglas P.
;  APPLICANT:  Taylor, Adriann J.
;  APPLICANT:  Parker, Margaret R.
;  TITLE OF INVENTION: METHODS AND COMPOSITIONS FOR THE
;  TITLE OF INVENTION:  DETECTION OF CERVICAL DISEASE
;  FILE REFERENCE: 046143/287139
;  CURRENT APPLICATION NUMBER: US/11/087,227
;  CURRENT FILING DATE:  2005-03-23
;  PRIOR APPLICATION NUMBER: 60/556,495
;  PRIOR FILING DATE: 2004-03-24
;  NUMBER OF SEQ ID NOS: 90
;  SOFTWARE: FastSEQ for Windows Version 4.0
; SEQ ID NO 13
;   LENGTH: 2821
;   TYPE: DNA
;   ORGANISM: Homo sapiens
US-11-087-227-13

Alignment Scores:
Pred. No.:               3.33e-62      Length:          2821
Score:                   1016.00       Matches:         298
Percent Similarity:      46.88%        Conservative:    160
Best Local Similarity:   30.50%        Mismatches:      324
Query Match:             22.04%        Indels:          195
DB:                      7             Gaps:            34
```

US-10-768-511-8 (1-901) x US-11-087-227-13 (1-2821)

```
Qy     26 ProLeuProGlnValThrSerProSerPheAspAsnAlaAlaSerProValAlaGlyArg 45
Db     10 CCAGCCCCAAGGGTCTAGGATACAGTCTTTGTAGATGAGCGGGTCCCCCTTGGAGGACAG 69

Qy     46 ArgAlaValArgGlnThrProThrSerAlaValArgArgArgGlyArgGluThrAspSer 65
Db     70 --------------------AATGAAGAATTGGGAAATCATGGCCGTTCTGGAGAGTAG 108

Qy     66 AlaArgArgArgArg-----------------------------SerArgSerArgSer 75
Db    109 ACAAGAAGACGGCGAAAGTCGGGCCTGCCCCGCCCTGCGGCCCCGGAACAAAAGAACGCG 168

Qy     76 LeuGlyAsnSerVal---------TyrSerSerProTyrAspAlaGlyThrProGlyThr 92
Db    169 TGTGCGCTGGCCCTTTAAGAGCGATTCTCCTCCGCCCGCGCCAGCTCGGACCGCGGGAAA 228

Qy     93 ProGlyThrProValAlaThrProValTyrAlaThrProValGlyThrProMetGlyThr 112
Db    229 CCCGGCGCCTGCACTACCCCGCCCGGAGATTCCCTTCCGACGCCCGCACCGCCTCCCCGT 288

Qy    113 ProSerPheHisArgGlyThr----------------------------------------- 119
Db    289 CACTCATTCTAGGCCCGCACGGTGATTGGCTTGCGGCTAGCGCGAGGTGAAGAAGGCCGC 348

Qy    120 ------------------ProGlnTyrLysGlnArgSerGluLeuGly------------ 129
Db    349 CTTGTCCGATTGGCCCGCACGCAGTGGCGC---CGGTCACGTGGGGGGCGACGTTTCGCG 405

Qy    130 ------SerGlnGlyLysProLeuHisArgArgArgArgSerGlnSerArgGluProGly 147
Db    406 CCAATTTCGGTTGGCCGGCCACAGTCCACCGCGCGGAGATTC---TCAGCTTCCCCAGGA 462

Qy    148 HisArgSerProSerArgGluProSerAlaAspGlyArgProSerGluSer---AlaGlu 166
Db    463 GCAAGA---CCTCTGAGCCCGCCAAGCGCGGCCGCACGGCCCTCGGCAGCGATGGCACTG 519

Qy    167 ProAspAspThrLeuGlyGlyGluTyrAlaTyrValTrpGlyThrAsnValAsnIlePro 186
Db    520 AAGGACTACGCGCTAGAGAAGGAA------------------------------------ 543

Qy    187 AspValLeuArgAlaIleArgArgPheLeuHisAsnTyrArgSerSerAlaHisAspLeu 206
Db    544 ------------AAGGTTAAGAAGTTCTTACAAGAGTTCTACCAGGAT---GATGAACTC 588

Qy    207 AsnSerLysTyrIleGlnIleIleGluGluThrVal---------GluArgGluGluAsp 223
Db    589 GGGAAGAAGCAGTTCAAGTATGGGAACCAGTTGGTTCGGCTGGCTCATCGGGAACAGGTG 648

Qy    224 ThrLeuAsnIleAspMetSerAspIleTyrAspHisAspProAspLeu------------ 239
Db    649 GCTCTGTATGTGGACCTGGACGACGTAGCCGAGGATGACCCCGAGTTGGTGGACTCAATT 708

Qy    240 ------------------TyrAlaLysIleValArgTyrProLeu---AspIleIlePro 252
Db    709 TGTGAGAATGCCAGGCGCTACGCGAAGCTCTTTGCTGATGCCGTACAAGAGCTGCTGCCT 768

Qy    253 ---------------------------------LeuLeuAspThrGluCysGln------ 259
Db    769 CAGTACAAGGAGAGGGAAGTGGTAAATAAAGATGTCCTGGACGTTTACATTGAGCATCGG 828

Qy    260 ------------------------------------------------------GluVal 261
Db    829 CTAATGATGGAGCAGCGGAGTCGGGACCCTGGGATGGTCCGAAGCCCCCAGAACCAGTAC 888

Qy    262 AlaThrSerLeuLeuProThrPheGluLysHisIleGluAlaArgProPheAsnLeuLys 281
Db    889 CCTGCTGAACTCATGCGCAGATTTGAGCTGTATTTTCAAGGC---CCTAGCAGCAACAAG 945

Qy    282 AlaSerValHisMetArgGluLeuAsnProSerAspIleAspLysLeuValSerValLys 301
Db    946 CCTCGTGTG---ATCCGGGAAGTGCGGGCTGACTCTGTGGGAAGTTGGTAACTGTGCGT 1002

Qy    302 GlyMetValIleArgCysSerSerIleIleProGluIleLysGlyAlaPhePheLysCys 321
Db   1003 GGAATCGTCACTCGTGTCTCTGAAGTCAAACCCAAGATGGTGGTGGCCACTTACACTTGT 1062

Qy    322 LeuValCysGly---------HisSerProProLeuValThrValValLysGlyArgVal 338
Db   1063 GACCAGTGTGGGGCAGAGACCTACCAGCCG-------------------------ATC 1095

Qy    339 GluGluProThr---------------ArgCysGluLysProGluCysAlaAlaArgAsn 353
Db   1096 CAGTCTCCCACTTTCATGCCTCTGATCATGTGCCCAAGCCAGGAGTGCCAAACCAACCGC 1155

Qy    354 Ala---------MetSerLeuIleHisAsnArgCysThrPheAlaAsnLysGlnIleVal 370
Db   1156 TCAGGAGGGCGGCTGTATCTGCAGACACGGGGCTCCAGATTCATCAAATTCCAGGAGATG 1215

Qy    371 ArgLeuGlnGluThrProAspAlaIleProGluGlyGluThrProHisThrValSerMet 390
Db   1216 AAGATGCAAGAACATAGTGATCAGGTGCCTGTGGGAAATATCCCTCGTAGTATCACGGTG 1275

Qy    391 CysLeuTyrAsnThrMetValAspAlaValLysProGlyAspArgIleGluValThrGly 410
Db   1276 CTGGTAGAAGGAGAGAACACAAGGATTGCCCAGCCTGGAGACCACGTCAGCGTCACTGGT 1335

Qy    411 ValPheLysAlaMetAlaValArgValGlyProAsnGlnArgThrLeuArgAlaLeuTyr 430
Db   1336 ATTTTCTTGCCAATC---CTGCGCACTGGGTTCCGACAGGTGGTACAGGGTTTACTCTCA 1392

Qy    431 LysThrTyrIleAspCysValHisValLysLysSerAspArgGlyArgLeuGlnThrGlu 450
Db   1393 GAAACCTACCTGGAAGCCCATCGGATTGTGAAG------------------------- 1425

Qy    451 AspProMetGluMetAspLysGluAsnAspMetTyrAlaGlyTyrHisGluSerAspThr 470
Db   1426 ------------ATGAACAAGAGTGAGGATGATGAGTCTGGG----------------- 1455

Qy    471 SerGluAlaAlaAsnGluAlaLysIleGlnLysLeuLysGluLeuSerLysLeuProGly 490
Db   1456 ---------GCTGGAGAGCTCACCAGGGAGGAGCTGAGGCAAATTGCAGAG---GAGGAT 1503

Qy    491 IleTyrAspArgLeuSerArgSerLeuAlaProSerIleTrpGluLeuGluAspIleLys 510
Db   1504 TTCTACGAAAAGCTGGCAGCTTCAATCGCCCCAGAAATATACGGGCATGAAGATGTGAAG 1563

Qy    511 LysGlyLeuLeuCysGlnLeuPheGlyGlyLysAlaLysLysIleProSerGlyAlaSer 530
Db   1564 AAGGCACTGCTGCTCCTGCTAGTCGGGGGT---GTGGACCAGTCTCCTCGAGGCATGAAA 1620

Qy    531 PheArgGlyAspIleAsnValLeuLeuValGlyAspProGlyThrSerLysSerGlnLeu 550
Db   1621 ATCCGGGGCAACATCAACATCTGTCTGATGGGGGATCCTGGTGTGGCCAAGTCTCAGCTC 1680

Qy    551 LeuGlnTyrValHisLysIleAlaProArgGlyIleTyrThrSerGlyArgGlySerSer 570
Db   1681 CTGTCATACATTGATCGACTGGCGCCTCGCAGCCAGTACACAACAGGCCGGGGCTCCTCA 1740

Qy    571 AlaValGlyLeuThrAlaTyrValThrLysAspProGluThrArgGluThrValLeuGlu 590
Db   1741 GGAGTGGGGCTTACGGCAGCTGTGCTGAGAGACTCCGTGAGTGGAGAACTGACCTTAGAG 1800

Qy    591 SerGlyAlaLeuValLeuSerAspArgGlyIleCysCysIleAspGluPheAspLysMet 610
Db   1801 GGTGGGGCCCTGGTGCTGGCTGACCAGGGTGTGTGCTGCATTGATGAGTTCGACAAGATG 1860

Qy    611 SerAspAsnAlaArgSerMetLeuHisGluValMetGluGlnGlnThrValSerValAla 630
Db   1861 GCTGAGGCCGACCGCACAGCCATCCACGAGGTCATGGAGCAGCAGACCATCTCCATTGCC 1920

Qy    631 LysGlyGlyIleIleAlaSerLeuAsnAlaArgThrSerValLeuAlaCysAlaAsnPro 650
Db   1921 AAGGCCGGCATTCTCACCACACTCAATGCCCGCTGCTCCATCCTGGCTGCCGCCAACCCT 1980
```

```
Qy    651 SerGlySerArgTyrAsnAlaArgLeuSerValIleAspAsnIleGlnLeuProProThr 670
Db   1981 GCCTACGGGCGCTACAACCCTCGCCGCAGCCTGGAGCAGAACATACAGCTACCTGCTGCA 2040

Qy    671 LeuLeuSerArgPheAspLeuIleTyrLeuMetLeuAspLysProAspGluGlnAsnAsp 690
Db   2041 CTGCTCTCCCGGTTTGACCTCCTCTGGCTGATTCAGGACCGGCCCGACCGAGACAATGAC 2100

Qy    691 ArgArgLeuAlaArgHisLeuValAlaLeuHisTyrGluAsnTyrGluValSerLysGln 710
Db   2101 CTACGGTTGGCCCAGCACATCACCTATGTGCACCAGCACAGCCGGCAGCCCCCCTCCCAG 2160

Qy    711 ---AspAlaLeuAspLeuGlnThrLeuThrAlaTyrIleThrTyrAlaArgGlnHisVal 729
Db   2161 TTTGAACCTCTGGACATGAAGCTCATGAGGCGTTACATAGCCATGTGCCGCGAGAAG--- 2217

Qy    730 HisProThrLeuSerAspGluAlaAlaGluAspLeuIleAsnGlyTyrValGluMetArg 749
Db   2218 CAGCCCATGGTGCCAGAGTCTCTGGCTGACTACATCACAGCAGCATACGTGGAGATGAGG 2277

Qy    750 GlnLysGlyAsnPheProGlySerSerLysLysValIleThrAlaThrProArgGlnLeu 769
Db   2278 CGAGAGGCT---------TGGGCTAGTAAGGATGCCACCTATACTTCTGCCCGGACCCTG 2328

Qy    770 GluSerMetIleArgIleSerGluAlaLeuAlaArgMetArgPheSerGluValValGlu 789
Db   2329 CTGGCTATCCTGCGCCTTTCCACTGCTCTGGCACGTCTGAGAATGGTGGATGTGGTGGAG 2388

Qy    790 LysValAspAlaAlaGluAlaValArgLeuLeuAspValAlaLeuGlnGlnSerAlaThr 809
Db   2389 AAAGAAGATGTGAATGAAGCCATCAGGCTAATGGAGATGTCAAAGGACTCTCTTCTAGGA 2448

Qy    810 AspHisAlaThrGlyThrIleAspMetAspLeuIleThrThrGlyValSerAlaSerGlu 829
Db   2449 GAC-----------------------------------AAGGGGCAGACAGCTAGGACT 2472

Qy    830 ArgIleArgArgAlaAsnLeuLeuAlaAlaLeuArgGluLeuIleAlaAspLysIleSer 849
Db   2473 CAGAGACCAGCAGATGTGATATTTGCCACCGTCCGTGAACTG------------GTCTCA 2520

Qy    850 ProGlySerSer-SerGlyLeuLysThrSerGlnLeuLeuGluAspIleArgSerGlnSe 869
Db   2521 GGGGGCCGAAGTGTCCGGT-----------------------TCTCTGAGGCAGAGCAG 2556

Qy    869 rSerValAspVal--------------------SerLeuGlnAspIleLysAsnAlaLe 882
Db   2557 CGCTGTGTATCTCGTGGCTTCACACCCGCCCAGTTCCAGGCGGCTCTGGATGAATATGAG 2616

Qy    882 uGlySerLeuGlnGlyGluGlyPheLeuThrValHisGlyAspIleVal 898
Db   2617 GAGCTCAATGTCTGGCAGGTCAATGCTTCCCGGACACGGATCACTTTTG 2665
```

```
RESULT 2
US-11-087-227-15
; Sequence 15, Application US/11087227
; Publication No. US20050260566A1
; GENERAL INFORMATION:
;  APPLICANT: Fischer, Timothy J.
;  APPLICANT:  Malinowski, Douglas P.
;  APPLICANT:  Taylor, Adriann J.
;  APPLICANT:  Parker, Margaret R.
;  TITLE OF INVENTION: METHODS AND COMPOSITIONS FOR THE
;  TITLE OF INVENTION:  DETECTION OF CERVICAL DISEASE
;  FILE REFERENCE: 046143/287139
;  CURRENT APPLICATION NUMBER: US/11/087,227
;  CURRENT FILING DATE:  2005-03-23
;  PRIOR APPLICATION NUMBER: 60/556,495
;  PRIOR FILING DATE: 2004-03-24
;  NUMBER OF SEQ ID NOS: 90
;  SOFTWARE: FastSEQ for Windows Version 4.0
; SEQ ID NO 15
;   LENGTH: 2900
;   TYPE: DNA
;   ORGANISM: Homo sapiens
US-11-087-227-15

Alignment Scores:
Pred. No.:              3.25e-59      Length:         2900
Score:                  974.00        Matches:        295
Percent Similarity:     47.42%        Conservative:   164
Best Local Similarity:  30.48%        Mismatches:     344
Query Match:            21.13%        Indels:         165
DB:                     7             Gaps:           32

US-10-768-511-8 (1-901) x US-11-087-227-15 (1-2900)
```

```
Qy      6 AlaLeuAspIleGlyAlaValSerSerProTyrProSerGlnSerGluGlyValSerThr 25
Db    121 GCGGTGTCTCTGGGTGTGATGGCCAATGGCTGGACTGGCTCCCGCCCTGGGCGGAGGAAT 180

Qy     26 Pro-----------------LeuProGlnValThrSerProSerPheAspAsnAlaAla 39
Db    181 CCCGAGCTGTGAAGCGGCTGGAATCCGGGCCCATGTGCTTCTTTGTTTACTAAGAGCGGA 240

Qy     40 SerPro-----------------ValAlaGlyArgArgAlaValArgGlnThrProThr 53
Db    241 AGCGATGGCGGGAGCGGGGGTGGGGTGCGGTGGCGGGGTGCGGTGGCGGAGGTCCCGGTG 300

Qy     54 SerAla--------------ValArgArgArgGly--------------ArgGluThr 63
Db    301 AAATCAGGGGCTAAGGGGACCCAAAGAAGGCGGGGGATCATAGGGGTGGAAAGAAAGCTG 360

Qy     64 AspSerAlaArgArgArgArgSerArgSerArgSerLeuGlyAsnSerValTyrSerSer 83
Db    361 AGAACCTTGAGACCGGAGTGTGAGGGGCCAACGGGGAAGGGCGCTAGAATTTTAAACTAA 420

Qy     84 ProTyrAspAlaGlyThrPro--GlyThrProGlyThrProValAlaThrPro------V 101
Db    421 AGTAGGGACCGGAATTCCCCTGGGGAGATGTTGGATGGCCCTGTGCACTGCCACGGGCTC 480

Qy    101 alTyrAlaThrProValGlyThrProMetGlyThrProSerPhe---HisArgGlyThrP 120
Db    481 TTTATTCTTCGCTGGTTAGAAACAGACTTGTGAAAAAGAGTTATGCCCACTTTGGGGAGA 540

Qy    120 roGlnTyrLysGlnArgSerGluLeuGlySerGlnGlyLysProLeuHisArgArgArgA 140
Db    541 CTTCGAAAAGGTTAAGAAGTTCTTACAAGAGTTCTACCAGGATGATGAACTCGGGAAGAA 600

Qy    140 rgSerGlnSerArg-GluPro---GlyHisArgSerProSerArgGluProSerAlaAsp 158
Db    601 GCAGTTCAAGTATGGGAACCAGTTGGTTCGGCTGGCTCATCGGGAACAGGTGGCTCTGTA 660

Qy    159 GlyArgPro-SerGluSerAlaGluProAspAspThrLeuGlyGlyGluTyrAlaTyrVa 178
Db    661 TGTGGACCTGGACGACGTAGCCGAGGATGACCCCGAGTTG------------------- 700

Qy    178 lTrpGlyThrAsnValAsnIleProAspValLeuArgAlaIleArgArgPheLeuHisAs 198
Db    701 ------------------GTGGACTCAATTTGTGAGAATGCCAGGCGCTACGCGAAGCT 741

Qy    198 nTyrArgSerSerAlaHisAspLeuAsnSerLysTyrIleGlnIleIleGluGluThrVa 218
Db    742 CTTTGCTGATGCCGTACAAGAGCTGCTGCCTCAGTAC---------------------- 778

Qy    218 lGluArgGluGluAspThrLeuAsnIleAspMetSerAspIleTyr------------- 233
Db    779 ----AAGGAGAGGGAAGTGGTAAATAAAGATGTCCTGGACGTTTACATTGAGCATCGGCT 834

Qy    234 ---------------AspHisAspProAspLeuTyrAlaLysIleValArg-------- 245
Db    835 AATGATGGAGCAGCGGAGTCGGGACCCTGGGATG-----------GTCCGAAGCCCCCA 882

Qy    246 -------TyrProLeuAspIleIleProLeuLeu-------------------AspTh 256
Db    883 GAACCAGTACCCTGCTGAACTCATGCGCAGATTGTGAGTGGTCTCTGTCGGGAAAGATGT 942
```

```
Qy        256 rGluCysGlnGluValAlaThrSerLeu---------LeuProThrPheGluLysHisIl 273
Db        943 AGGGATTGGTTCTCCAGGATCTTGTTTGTGACTGTTTTCTCCCCTTAGTGAGCTGTATTT 1002

Qy        273 eGluAlaArgProPheAsnLeuLysAlaSerValHisMetArgGluLeuAsnProSerAs 293
Db       1003 TCAAGGC---CCTAGCAGCAACAAGCCTCGTGTG---ATCCGGGAAGTGCGGGCTGACTC 1056

Qy        293 pIleAspLysLeuValSerValLysGlyMetValIleArgCysSerSerIleIleProGl 313
Db       1057 TGTGGGGAAGTTGGTAACTGTGCGTGGAATCGTCACTCGTGTCTCTGAAGTCAAACCCAA 1116

Qy        313 uIleLysGlyAlaPhePheLysCysLeuValCysGly---------HisSerProProLe 330
Db       1117 GATGGTGGTGGCCACTTACACTTGTGACCAGTGTGGGGCAGAGACCTACCAGCCG----- 1171

Qy        330 uValThrValValLysGlyArgValGluGluProThr---------------ArgCysGl 345
Db       1172 ---------------------ATCCAGTCTCCCACTTTCATGCCTCTGATCATGTGCCC 1209

Qy        345 uLysProGluCysAlaAlaArgAsnAla---------MetSerLeuIleHisAsnArgCy 362
Db       1210 AAGCCAGGAGTGCCAAACCAACCGCTCAGGAGGGCGGCTGTATCTGCAGACACGGGGCTC 1269

Qy        362 sThrPheAlaAsnLysGlnIleValArgLeuGlnGluThrProAspAlaIleProGluGl 382
Db       1270 CAGATTCATCAAATTCCAGGAGATGAAGATGCAAGAACATAGTGATCAGGTGCCTGTGGG 1329

Qy        382 yGluThrProHisThrValSerMetCysLeuTyrAsnThrMetValAspAlaValLysPr 402
Db       1330 AAATATCCCTCGTAGTATCACGGTGCTGGTAGAAGGAGAGAACACAAGGATTGCCCAGCC 1389

Qy        402 oGlyAspArgIleGluValThrGlyValPheLysAlaMetAlaValArgValGlyProAs 422
Db       1390 TGGAGACCACGTCAGCGTCACTGGTATTTTCTTGCCAATC---CTGCGCACTGGGTTCCG 1446

Qy        422 nGlnArgThrLeuArgAlaLeuTyrLysThrTyrIleAspCysValHisValLysLysSe 442
Db       1447 ACAGGTGGTACAGGGTTTACTCTCAGAAACCTACCTGGAAGCCCATCGGATTGTGAAG-- 1504

Qy        442 rAspArgGlyArgLeuGlnThrGluAspProMetGluMetAspLysGluAsnAspMetTy 462
Db       1505 ------------------------------------ATGAACAAGAGTGAGGATGATGA 1527

Qy        462 rAlaGlyTyrHisGluSerAspThrSerGluAlaAlaAsnGluAlaLysIleGlnLysLe 482
Db       1528 GTCTGGG--------------------------GCTGGAGAGCTCACCAGGGAGGAGCT 1560

Qy        482 uLysGluLeuSerLysLeuProGlyIleTyrAspArgLeuSerArgSerLeuAlaProSe 502
Db       1561 GAGGCAAATTGCAGAG---GAGGATTTCTACGAAAAGCTGGCAGCTTCAATCGCCCCAGA 1617

Qy        502 rIleTrpGluLeuGluAspIleLysLysGlyLeuLeuCysGlnLeuPheGlyGlyLysAl 522
Db       1618 AATATACGGGCATGAAGATGTGAAGAAGGCACTGCTGCTCCTGCTAGTCGGGGGT---GT 1674

Qy        522 aLysLysIleProSerGlyAlaSerPheArgGlyAspIleAsnValLeuLeuValGlyAs 542
Db       1675 GGACCAGTCTCCTCGAGGCATGAAAATCCGGGGCAACATCAACATCTGTCTGATGGGGGA 1734

Qy        542 pProGlyThrSerLysSerGlnLeuLeuGlnTyrValHisLysIleAlaProArgGlyIl 562
Db       1735 TCCTGGTGTGGCCAAGTCTCAGCTCCTGTCATACATTGATCGACTGGCGCCTCGCAGCCA 1794

Qy        562 eTyrThrSerGlyArgGlySerSerAlaValGlyLeuThrAlaTyrValThrLysAspPr 582
Db       1795 GTACACAACAGGCCGGGGCTCCTCAGGAGTGGGGCTTACGGCAGCTGTGCTGAGAGACTC 1854

Qy        582 oGluThrArgGluThrValLeuGluSerGlyAlaLeuValLeuSerAspArgGlyIleCy 602
Db       1855 CGTGAGTGGAGAACTGACCTTAGAGGGTGGGGCCCTGGTGCTGGCTGACCAGGGTGTGTG 1914

Qy        602 sCysIleAspGluPheAspLysMetSerAspAsnAlaArgSerMetLeuHisGluValMe 622
Db       1915 CTGCATTGATGAGTTCGACAAGATGGCTGAGGCCGACCGCACAGCCATCCACGAGGTCAT 1974

Qy        622 tGluGlnGlnThrValSerValAlaLysGlyGlyIleIleAlaSerLeuAsnAlaArgTh 642
Db       1975 GGAGCAGCAGACCATCTCCATTGCCAAGGCCGGCATTCTCACCACACTCAATGCCCGCTG 2034

Qy        642 rSerValLeuAlaCysAlaAsnProSerGlySerArgTyrAsnAlaArgLeuSerValIl 662
Db       2035 CTCCATCCTGGCTGCCGCCAACCCTGCCTACGGGCGCTACAACCCTCGCCGCAGCCTGGA 2094

Qy        662 eAspAsnIleGlnLeuProProThrLeuLeuSerArgPheAspLeuIleTyrLeuMetLe 682
Db       2095 GCAGAACATACAGCTACCTGCTGCACTGCTCTCCCGGTTTGACCTCCTCTGGCTGATTCA 2154

Qy        682 uAspLysProAspGluGlnAsnAspArgArgLeuAlaArgHisLeuValAlaLeuHisTy 702
Db       2155 GGACCGGCCCGACCGAGACAATGACCTACGGTTGGCCCAGCACATCACCTATGTGCACCA 2214

Qy        702 rGluAsnTyrGluValSerLysGln---AspAlaLeuAspLeuGlnThrLeuThrAlaTy 721
Db       2215 GCACAGCCGGCAGCCCCCCTCCCAGTTTGAACCTCTGGACATGAAGCTCATGAGGCGTTA 2274

Qy        721 rIleThrTyrAlaArgGlnHisValHisProThrLeuSerAspGluAlaAlaGluAspLe 741
Db       2275 CATAGCCATGTGCCGCGAGAAG---CAGCCCATGGTGCCAGAGTCTCTGGCTGACTACAT 2331

Qy        741 uIleAsnGlyTyrValGluMetArgGlnLysGlyAsnPheProGlySerSerLysLysVa 761
Db       2332 CACAGCAGCATACGTGGAGATGAGGCGAGAGGCT---------TGGGCTAGTAAGGATGC 2382

Qy        761 lIleThrAlaThrProArgGlnLeuGluSerMetIleArgIleSerGluAlaLeuAlaAr 781
Db       2383 CACCTATACTTCTGCCCGGACCCTGCTGGCTATCCTGCGCCTTTCCACTGCTCTGGCACG 2442

Qy        781 gMetArgPheSerGluValValGluLysValAspAlaAlaGluAlaValArgLeuLeuAs 801
Db       2443 TCTGAGAATGGTGGATGTGGTGGAGAAAGAAGATGTGAATGAAGCCATCAGGCTAATGGA 2502

Qy        801 pValAlaLeuGlnGlnSerAlaThrAspHisAlaThrGlyThrIleAspMetAspLeuIl 821
Db       2503 GATGTCAAAGGACTCTCTTCTAGGAGAC------------------------------- 2530

Qy        821 eThrThrGlyValSerAlaSerGluArgIleArgArgAlaAsnLeuLeuAlaAlaLeuAr 841
Db       2531 ----AAGGGGCAGACAGCTAGGACTCAGAGACCAGCAGATGTGATATTTGCCACCGTCCG 2586

Qy        841 gGluLeuIleAlaAspLysIleSerProGlySerSer-SerGlyLeuLysThrSerGlnL 861
Db       2587 TGAACTG------------GTCTCAGGGGGCCGAAGTGTCCGGT---------------- 2618

Qy        861 euLeuGluAspIleArgSerGlnSerSerValAspVal---------------------S 874
Db       2619 --------TCTCTGAGGCAGAGCAGCGCTGTGTATCTCGTGGCTTCACACCCGCCCAGTT 2670

Qy        874 erLeuGlnAspIleLysAsnAlaLeuGlySerLeuGlnGlyGluGlyPheLeuThrValH 894
Db       2671 CCAGGCGGCTCTGGATGAATATGAGGAGCTCAATGTCTGGCAGGTCAATGCTTCCCGGAC 2730

Qy        894 isGlyAspIleVal 898
Db       2731 ACGGATCACTTTTG 2744
```

RESULT 3
US-11-087-227-11
; Sequence 11, Application US/11087227
; Publication No. US20050260566A1
; GENERAL INFORMATION:
; APPLICANT: Fischer, Timothy J.
; APPLICANT: Malinowski, Douglas P.
; APPLICANT: Taylor, Adriann J.
; APPLICANT: Parker, Margaret R.
; TITLE OF INVENTION: METHODS AND COMPOSITIONS FOR THE

```
;  TITLE OF INVENTION:  DETECTION OF CERVICAL DISEASE
;  FILE REFERENCE: 046143/287139
;  CURRENT APPLICATION NUMBER: US/11/087,227
;  CURRENT FILING DATE:  2005-03-23
;  PRIOR APPLICATION NUMBER: 60/556,495
;  PRIOR FILING DATE: 2004-03-24
;  NUMBER OF SEQ ID NOS: 90
;  SOFTWARE: FastSEQ for Windows Version 4.0
; SEQ ID NO 11
;   LENGTH: 3453
;   TYPE: DNA
;   ORGANISM: Homo sapiens
US-11-087-227-11

Alignment Scores:
Pred. No.:              5.84e-57     Length:        3453
Score:                  943.50       Matches:       261
Percent Similarity:     46.44%       Conservative:  157
Best Local Similarity:  29.00%       Mismatches:    343
Query Match:            20.47%       Indels:        139
DB:                     7            Gaps:          23

US-10-768-511-8 (1-901) x US-11-087-227-11 (1-3453)

Qy      31 ThrSerProSerPheAspAsnAlaAlaSerProValAlaGlyArgArgAlaValArgGln 50
Db      67 TCATCGGAATCCTTCACCATGGCATCCAGCCCG---GCCCAGCGTCGGCGAGGCAATGAT 123
Qy      51 ThrProThrSerAlaValArgArgArgGlyArgGluThrAspSerAlaArgArgArgArg 70
Db     124 CCTCTCACCTCCAGCCCTGGCCGAAGCTCCCGGCGTACTGATGCCCTC------ACCTCC 177
Qy      71 SerArgSerArgSerLeuGlyAsnSerValTyrSerSerProTyrAspAlaGlyThrPro 90
Db     178 AGCCCTGGCCGTGACCTT------------------CCACCATTTGAGGATGAGTCCGAG 219
Qy      91 GlyThrProGlyThrProValAlaThrProValTyrAlaThrProValGlyThrProMet 110
Db     220 GGGCTCCTAGGCACA------GAGGGGCCCCTGGAGGAAGAAGAGGATGGAGAGGAGCTC 273
Qy     111 ---------GlyThrProSerPheHisArgGlyThrProGlnTyrLysGlnArgSerGlu 127
Db     274 ATTGGAGATGGCATGGAAAGGGACTACCGCGCCATCCCAGAGCTGGACGCCTATGAGGCC 333
Qy     128 LeuGly---------------------------------------SerGlnGlyLysPro 134
Db     334 GAGGGACTGGCTCTGGATGATGAGGACGTAGAGGAGCTGACGGCCAGTCAGAGGGAGGCA 393
Qy     135 LeuHisArgArgArgArgSerGlnSerArgGluProGly--------------------- 147
Db     394 GCAGAGCGGGCCATGCGGCAGCGTGACCGGGAGGCTGGCCGGGGCCTGGGCCGCATGCGC 453
Qy     148 ---------------------------------HisArgSerProSerArgGluProSer 156
Db     454 CGTGGGCTCCTGTATGACAGCGATGAGGAGGACGAGGAGCGCCCTGCCCGCAAGCGCCGC 513
Qy     157 AlaAspGlyArgProSerGluSerAlaGluProAsp------------------------ 168
Db     514 CAGGTGGAGCGGGCCACGGAGGACGGCGAGGAGGACGAGGAGATGATCGAGAGCATCGAG 573
Qy     169 ------AspThrLeuGlyGlyGluTyrAlaTyrValTrp------------GlyThrAsn 182
Db     574 AACCTGGAGGATCTCAAAGGCCACTCTGTGCGCGAGTGGGTGAGCATGGCGGGCCCCCGG 633
Qy     183 ValAsnIleProAspValLeuArgAlaIleArgArgPheLeuHisAsnTyr---ArgSer 201
Db     634 CTG---------GAGATCCACCACCGCTTCAAGAACTTCCTGCGCACTCACGTCGACAGC 684
Qy     202 SerAlaHisAspLeuAsnSerLysTyrIleGlnIleIleGluGluThrValGluArgGlu 221
Db     685 CACGGCCACAACGTC---------TTCAAGGAGCGCATCAGCGACATGTGCAAAGAGAAC 735
Qy     222 GluAspThrLeuAsnIleAspMetSerAspIleTyrAspHisAspProAspLeuTyrAla 241
Db     736 CGTGAGAGCCTGGTGGTGAACTATGAGGACTTGGCAGCCAGGGAGCACGTGCTGGCCTAC 795
Qy     242 LysIleValArgTyrProLeuAspIleIleProLeuLeuAspThrGluCysGlnGluVal 261
Db     796 TTCCTGCCTGAGGCACCGGCGGAGCTGCTGCAGATCTTTGATGAGGCTGCCCTGGAGGTG 855
Qy     262 AlaThrSerLeuLeuProThrPheGluLys---------HisIleGluAlaArgProPhe 278
Db     856 GTACTGGCCATGTACCCCAAGTACGACCGCATCACCAACCACATCCATGTCCGCATCTCC 915
Qy     279 AsnLeuLysAlaSerValHisMetArgGluLeuAsnProSerAspIleAspLysLeuVal 298
Db     916 CACCTGCCTCTGGTGGAGGAGCTGCGCTCGCTGAGGCAGCTGCATCTGAACCAGCTGATC 975
Qy     299 SerValLysGlyMetValIleArgCysSerSerIleIleProGluIleLysGlyAlaPhe 318
Db     976 CGCACCAGTGGGGTGGTGACCAGCTGCACTGGCGTCCTGCCCCAGCTCAGCATGGTCAAG 1035
Qy     319 PheLysCysLeuValCysGlyHisSerProProLeuValThrValValLysGlyArgVal 338
Db    1036 TACAACTGCAACAAGTGCAATTTCGTCCTGGGTCCTTTCTGCCAGTCCCAGAACCAGGAG 1095
Qy     339 GluGluProThrArgCysGluLysProGluCysAlaAlaArgAsnAlaMetSerLeuIle 358
Db    1096 GTGAAACCAGGCTCCTGT------CCTGAGTGCCAGTCGGCCGGCCCCTTTGAGGTCAAC 1149
Qy     359 HisAsnArgCysThrPheAlaAsnLysGlnIleValArgLeuGlnGluThrProAspAla 378
Db    1150 ATGGAGGAGACCATCTATCAGAACTACCAGCGTATCCGAATCCAGGAGAGTCCAGGCAAA 1209
Qy     379 IleProGluGlyGluThrProHisThrValSerMetCysLeuTyrAsnThrMetValAsp 398
Db    1210 GTGGCGGCTGGCCGGCTGCCCCGCTCCAAGGACGCCATTCTCCTCGCAGATCTGGTGGAC 1269
Qy     399 AlaValLysProGlyAspArgIleGluValThrGlyValPheLysAlaMetAlaValArg 418
Db    1270 AGCTGCAAGCCAGGAGACGAGATAGAGCTGACTGGCATCTATCACAAC---------AAC 1320
Qy     419 ValGlyProAsnGlnArgThrLeuArgAla------LeuTyrLysThrTyrIleAspCys 436
Db    1321 TATGATGGCTCCCTCAACACTGCCAATGGCTTCCCTGTCTTTGCCACTGTCATCCTAGCC 1380
Qy     437 ValHisValLysLysSerAspArg------------GlyArgLeuGlnThrGluAspPro 452
Db    1381 AACCACGTGGCCAAGAAGGACAACAAGGTTGCTGTAGGGGAACTGACCGATGAAGAT--- 1437
Qy     453 MetGluMetAspLysGluAsnAspMetTyrAlaGlyTyrHisGluSerAspThrSerGlu 472
Db    1437 ------------------------------------------------------------ 1437
Qy     473 AlaAlaAsnGluAlaLysIleGlnLysLeuLysGluLeuSerLysLeuProGlyIleTyr 492
Db    1438 -----------------GTGAAGATGATCACTAGCCTCTCCAAGGATCAGCAGATCGGA 1479
Qy     493 AspArgLeuSerArgSerLeuAlaProSerIleTrpGluLeuGluAspIleLysLysGly 512
Db    1480 GAGAAGATCTTTGCCAGCATTGCTCCTTCCATCTATGGTCATGAAGACATCAAGAGAGGC 1539
Qy     513 LeuLeuCysGlnLeuPheGlyGlyLysAlaLysLysIleProSerGlyAlaSerPheArg 532
Db    1540 CTGGCTCTGGCCCTGTTCGGAGGGGAGCCCAAAAACCCAGGTGGCAAGCACAAGGTACGT 1599
Qy     533 GlyAspIleAsnValLeuLeuValGlyAspProGlyThrSerLysSerGlnLeuLeuGln 552
Db    1600 GGTGATATCAACGTGCTCTTGTGCGGAGACCCTGGCACAGCGAAGTCGCAGTTTCTCAAG 1659
Qy     553 TyrValHisLysIleAlaProArgGlyIleTyrThrSerGlyArgGlySerSerAlaVal 572
Db    1660 TATATTGAGAAAGTGTCCAGCCGAGCCATCTTCACCACTGGCCAGGGGGCGTCGGCTGTG 1719
Qy     573 GlyLeuThrAlaTyrValThrLysAspProGluThrArgGluThrValLeuGluSerGly 592
```

```
Db       1720 GGCCTCACGGCGTATGTCCAGCGGCACCCTGTCAGCAGGGAGTGGACCTTGGAGGCTGGG 1779

Qy        593 AlaLeuValLeuSerAspArgGlyIleCysCysIleAspGluPheAspLysMetSerAsp 612
              |||||||||||:::||||||||:::|||  ||||||||| ||||||||||:::|||
Db       1780 GCCCTGGTTCTGGCTGACCGAGGAGTGTGTCTCATTGATGAATTTGACAAGATGAATGAC 1839

Qy        613 AsnAlaArgSerMetLeuHisGluValMetGluGlnGlnThrValSerValAlaLysGly 632
                    |||:::   :::|||||  |||||||||||:::::::|||::::::|||
Db       1840 CAGGACAGAACCAGCATCCATGAGGCCATGGAGCAACAGAGCATCTCCATCTCGAAGGCT 1899

Qy        633 GlyIleIleAlaSerLeuAsnAlaArgThrSerValLeuAlaCysAlaAsnProSerGly 652
              ||||||:::   ||||||  ||||||   :::|||:::|||  |||||||||   |||
Db       1900 GGCATCGTCACCTCCCTGCAGGCTCGCTGCACGGTCATTGCTGCCGCCAACCCCATAGGA 1959

Qy        653 SerArgTyrAsnAlaArgLeuSerValIleAspAsnIleGlnLeuProProThrLeuLeu 672
                 ||||||:::      |||:::      :::|||:::   |||       ::::::
Db       1960 GGGCGCTACGACCCCTCGCTGACTTTCTCTGAGAACGTGGACCTCACAGAGCCCATCATC 2019

Qy        673 SerArgPheAspLeuIleTyrLeuMetLeuAspLysProAspGluGlnAsnAspArgArg 692
              ||||||||||||::::::   ::::::   |||      |||         |||
Db       2020 TCACGCTTTGACATCCTGTGTGTGGTGAGGGACACCGTGGACCCAGTCCAGGACGAGATG 2079

Qy        693 LeuAlaArgHisLeuValAlaLeuHisTyrGluAsnTyrGluValSerLysGln------ 710
              |||||||||   :::|||    |||     ::::::     :::|||:::
Db       2080 CTGGCCCGCTTCGTGGTGGGCAGCCACGTCAGACACCACCCCAGCAACAAGGAGGAGGAG 2139

Qy        711 -----------------------------------------------------AspAla 712
                                                                   :::
Db       2140 GGGCTGGCCAATGGCAGCGCTGCTGAGCCCGCCATGCCCAACACGTATGGCGTGGAGCCC 2199

Qy        713 LeuAspLeuGlnThrLeuThrAlaTyrIleThrTyrAlaArgGlnHisValHisProThr 732
              |||      :::   |||      ||||||  ||||||::::::   |||||||||
Db       2200 CTGCCCCAGGAGGTCCTGAAGAAGTACATCATCTACGCCAAGGAGAGGGTCCACCCGAAG 2259

Qy        733 LeuSerAspGluAlaAlaGluAspLeuIleAsnGlyTyrValGluMetArgGlnLysGly 752
              |||:::         :::   :::         |||    ::::::|||::::::
Db       2260 CTCAACCAGATGGACCAGGACAAGGTGGCCAAGATGTACAGTGACCTGAGGAAAGAATCT 2319

Qy        753 AsnPheProGlySerSerLysLysValIleThrAlaThrProArgGlnLeuGluSerMet 772
                       ||||||         |||      |||   |||   :::|||||||||
Db       2320 ATGGCGACAGGCAGC------------ATCCCCATTACGGTGCGGCACATCGAGTCCATG 2367

Qy        773 IleArgIleSerGluAlaLeuAlaArgMetArgPheSerGluValValGluLysValAsp 792
              ||||||::::::||||||   ||||||:::      :::   |||   :::   |||
Db       2368 ATCCGCATGGCGGAGGCCCACGCGCGCATCCATCTGCGGGACTATGTGATCGAAGACGAC 2427

Qy        793 AlaAlaGluAlaValArgLeuLeuAspValAlaLeuGlnGlnSerAlaThrAspHisAla 812
                       |||:::|||:::            :::   :::|||      |||
Db       2428 GTCAACATGGCCATCCGCGTG------------ATGCTGGAGAGCTTCATAGACACACAG 2475

Qy        813 ThrGlyThrIleAspMetAspLeuIleThrThrGlyValSerAlaSerGluArgIleArg 832
                    ::::::         :::         |||      :::            |||
Db       2476 AAGTTCAGCGTCATGCGCAGCATGCGCAAGACT---TTTGCCCGCTACCTTTCATTCCGG 2532

Qy        833 ArgAlaAsn---------LeuLeuAlaAlaLeuArgGluLeuIleAlaAspLysIleSer 849
              |||   |||         ||||||      |||::::::|||:::|||::::::::::::
Db       2533 CGTGACAACAATGAGCTGTTGCTCTTCATACTGAAGCAGTTAGTGGCAGAGCAGGTGACA 2592
```

RESULT 4
US-10-507-275-4
; Sequence 4, Application US/10507275
; Publication No. US20050250166A1
; GENERAL INFORMATION:
; APPLICANT: Masai, Hisao
; APPLICANT: Tamai, Katsuyuki
; APPLICANT: Medical and Biological Laboratories Co., Ltd.
; APPLICANT: Japan Science and Technology Agency
; APPLICANT: Ginkgo Biomedical Research Institute Co., Ltd.
; TITLE OF INVENTION: Cdc7-ASK Kinase Complex, Substrates of the Kinase Complex,
; TITLE OF INVENTION: Specific Antibodies to the Substrates, and Screening Methods
; TITLE OF INVENTION: Using the Same to Screen for Compounds Comprising Cdc7-ASK
; TITLE OF INVENTION: Kinase Inhititory Ability
; FILE REFERENCE: 082368-001100US
; CURRENT APPLICATION NUMBER: US/10/507,275
; CURRENT FILING DATE: 2004-09-09
; PRIOR APPLICATION NUMBER: JP 2002-067702
; PRIOR FILING DATE: 2002-03-12
; PRIOR APPLICATION NUMBER: WO PCT/JP03/02918
; PRIOR FILING DATE: 2003-03-12
; NUMBER OF SEQ ID NOS: 21
; SOFTWARE: PatentIn Ver. 2.1
; SEQ ID NO 4
; LENGTH: 2715
; TYPE: DNA
; ORGANISM: Mus musculus
; FEATURE:
; NAME/KEY: CDS
; LOCATION: (1)..(2715)
US-10-507-275-4

Alignment Scores:

Pred. No.:	3.9e-56	Length:	2715
Score:	930.00	Matches:	258
Percent Similarity:	46.65%	Conservative:	160
Best Local Similarity:	28.79%	Mismatches:	336
Query Match:	20.18%	Indels:	142
DB:	6	Gaps:	23

US-10-768-511-8 (1-901) x US-10-507-275-4 (1-2715)

```
Qy         37 AsnAlaAlaSerProValAlaGlyArgArgAlaValArgGlnThrProThrSerAlaVal 56
              :::|||:::||||||                  ||||||   :::          ||||||:::
Db         25 TCAGCATCTAGCCCTGCCCGTCAGCGGCGCCGGATCAGTGATCCCCTCACCTCCAGCCCA 84

Qy         57 ArgArgArgGlyArgGluThrAspSerAlaArgArgArgArgSerArgSerArgSerLeu 76
                 |||                     |||:::||||||   :::   :::   |||
Db         85 GGCCGC------------------AGCTCCAGACGTGCTGACGCCCTGACCTCCAGCCCT 126

Qy         77 GlyAsnSerValTyrSerSerProTyrAspAlaGlyThrProGlyThrProGlyThrPro 96
                 |||      :::         |||::::::      :::   |||      ||||||
Db        127 GGCAGAGACCTC------CCCCCATTTGAAGATGAGTCTGAGGGGCTTCTGGGCACA--- 177

Qy         97 ValAlaThrProValTyrAlaThrProValGlyThrProMet---------GlyThrPro 113
                       |||:::                  |||      :::         |||
Db        178 ---GAGGGGCCCATGGAGGAAGAAGAGGATGGAGAGGAACTCATTGGTGATGGCATGGAG 234

Qy        114 SerPheHisArgGlyThrProGln------------------------------------ 121
                 :::|||      |||:::
Db        235 AGAGACTACCGTCCCATTCCGGAGCTCGATGTCTACGAGGCCGAGGGATTGGCCCTGGAT 294

Qy        122 ------------------------------------------------------TyrLys 123
                                                                    :::
Db        295 GATGAAGATGTGGAGGAGCTGACAGCCAGTCAGAGAGAGGCAGCTGAGCGGACCATGAGG 354

Qy        124 GlnArgSerGluLeuGlySerGlnGlyLysProLeuHisArgArgArgArgSer------ 141
                 ||||||                  |||:::   |||   |||   ||||||
Db        355 CAGCGGGACCGT------GAGGCTGGCAGAGGCCTGGGACGCATGCGCCGGGGGCTGCTC 408

Qy        142 ---GlnSerArgGluProGlyHisArgSerProSerArgGluProSerAlaAspGlyArg 160
                       |||   |||                  |||:::|||:::            |||
Db        409 TATGACAGCAGCGAGGAAGATGAGGAGCGGCCTGCCCGTAAGCGCCGCCACGTAGAACGC 468

Qy        161 ProSerGluSerAlaGluProAsp-----------------------------AspThr 170
                 :::|||      |||   |||                              :::
Db        469 GCCACAGAGGATGGCGAGGAGGATGAAGAGATGATCGAGAGTATTGAGAATCTGGAGGAC 528

Qy        171 LeuGlyGly------------GluTyrAlaTyrValTrpGlyThrAsnValAsnIlePro 186
              |||   |||                  |||         |||      :::
Db        529 CTCAAGGGCCACTCGGTGCGCGAGCGGGTGAGCATGGCAGGGCCCAGGCTG--------- 579

Qy        187 AspValLeuArgAlaIleArgArgPheLeuHisAsnTyr---ArgSerSerAlaHisAsp 205
              ::::::         :::   ||||||      :::      |||      |||:::
Db        580 GAGATCCACCACCGCTTCAAGAACTTCCTGCGCACCCACGTGGACAGCCATGGCCACAAC 639

Qy        206 LeuAsnSerLysTyrIleGlnIleIleGluGluThrValGluArgGluGluAspThrLeu 225
              :::         :::   :::   |||   :::      :::         ::::::|||
```

```
Db    640 GTC---------TTCAAGGAGCGCATCAGTGATATGTGCAAAGAGAACCGTGAGAGTTTG 690

Qy    226 AsnIleAspMetSerAspIleTyrAspHisAspProAspLeuTyrAlaLysIleValArg 245
Db    691 GTGGTAAATTATGAAGACCTGGCAGCCCGGGAGCACGTGTTGGCATACTTCCTGCCGGAA 750

Qy    246 TyrProLeuAspIleIleProLeuLeuAspThrGluCysGlnGluValAlaThrSerLeu 265
Db    751 GCACCGGCTGAGTTGCTGCAGATCTTTGACGAGGCTGCCCTGGAGGTCGTGTTGGCCATG 810

Qy    266 LeuProThrPheGluLys---------HisIleGluAlaArgProPheAsnLeuLysAla 282
Db    811 TACCCTAAATATGACCGTATCACCAACCACATCCATGTGCGCATCTCCCACCTGCCTCTG 870

Qy    283 SerValHisMetArgGluLeuAsnProSerAspIleAspLysLeuValSerValLysGly 302
Db    871 GTGGAGGAGCTGCGTTCACTGAGGCAGTTGCACCTGAACCAGCTGATCCGTACCAGTGGC 930

Qy    303 MetValIleArgCysSerSerIleIleProGluIleLysGlyAlaPhePheLysCysLeu 322
Db    931 GTGGTGACCAGCTGCACCGGAGTCCTGCCCCAGCTCAGCATGGTCAAGTACAACTGTAGC 990

Qy    323 ValCysGlyHisSerProProLeuValThrValValLysGlyArgValGluGluProThr 342
Db    991 AAGTGCAACTTTGTACTGGGGCCTTTCTGCCAGTCTCAGAATCAGGAGGTGAAGCCTGGC 1050

Qy    343 ArgCysGluLysProGluCysAlaAlaArgAsnAlaMetSerLeuIleHisAsnArgCys 362
Db   1051 TCCTGC------CCTGAGTGCCAGTCTGCTGGGCCCTTTGAGATCAACATGGAGGAGACC 1104

Qy    363 ThrPheAlaAsnLysGlnIleValArgLeuGlnGluThrProAspAlaIleProGluGly 382
Db   1105 ATCTATCAGAACTACCAACGTATCCGCATCCAGGAGAGTCCCGGCAAGGTGGCGGCTGGC 1164

Qy    383 GluThrProHisThrValSerMetCysLeuTyrAsnThrMetValAspAlaValLysPro 402
Db   1165 CGACTGCCCCGTTCCAAGGATGCCATTCTCCTCGCTGATCTGGTGGACAGCTGCAAGCCA 1224

Qy    403 GlyAspArgIleGluValThrGlyValPheLysAlaMetAlaValArgValGlyProAsn 422
Db   1225 GGGGACGAGATTGAGCTGACCGGCATTTACCATAAT---------AACTATGACGGCTCG 1275

Qy    423 GlnArgThrLeuArgAla------LeuTyrLysThrTyrIleAspCysValHisValLys 440
Db   1276 CTTAACACCGCCAACGGCTTTCCAGTCTTTGCCACTATTATCTTGGCCAACCATGTTGCC 1335

Qy    441 LysSerAspArg------------GlyArgLeuGlnThrGluAspProMetGluMetAsp 456
Db   1336 AAGAAGGACAACAAAGTAGCTGTGGGGGAGCTCACCGATGAGGAC--------------- 1380

Qy    457 LysGluAsnAspMetTyrAlaGlyTyrHisGluSerAspThrSerGluAlaAlaAsnGlu 476
Db   1380 ------------------------------------------------------------ 1380

Qy    477 AlaLysIleGlnLysLeuLysGluLeuSerLysLeuProGlyIleTyrAspArgLeuSer 496
Db   1381 ------GTGAAGATGATCACCGGTCTCTCCAAGGATCAGCAAATTGGAGAGAAGATCTTT 1434

Qy    497 ArgSerLeuAlaProSerIleTrpGluLeuGluAspIleLysLysGlyLeuLeuCysGln 516
Db   1435 GCCAGCATTGCACCCTCCATCTATGGGCATGAAGACATCAAGAGAGGCCTGGCTCTGGCC 1494

Qy    517 LeuPheGlyGlyLysAlaLysLysIleProSerGlyAlaSerPheArgGlyAspIleAsn 536
Db   1495 CTGTTTGGAGGGGAGCCCAAGAACCCAGGTGGAAAGCACAAGGTTCGAGGTGACATTAAT 1554

Qy    537 ValLeuLeuValGlyAspProGlyThrSerLysSerGlnLeuLeuGlnTyrValHisLys 556
Db   1555 GTGCTCTTGTGTGGGGACCCTGGCACAGCAAAGTCCCAATTCCTCAAATACATCGAGAAA 1614

Qy    557 IleAlaProArgGlyIleTyrThrSerGlyArgGlySerSerAlaValGlyLeuThrAla 576
Db   1615 GTGTCTAGCCGTGCCATCTTCACCACTGGCCAGGGTGCGTCAGCAGTGGGTCTCACCGCG 1674

Qy    577 TyrValThrLysAspProGluThrArgGluThrValLeuGluSerGlyAlaLeuValLeu 596
Db   1675 TACGTTCAGCGGCATCCCGTCAGCAGAGAGTGGACCTTAGAGGCGGGAGCCCTGGTTCTG 1734

Qy    597 SerAspArgGlyIleCysCysIleAspGluPheAspLysMetSerAspAsnAlaArgSer 616
Db   1735 GCTGACCGGGGGGTGTGTCTCATTGACGAGTTTGACAAGATGAATGACCAGGACAGGACC 1794

Qy    617 MetLeuHisGluValMetGluGlnGlnThrValSerValAlaLysGlyGlyIleIleAla 636
Db   1795 AGCATCCACGAGGCCATGGAACAGCAAAGCATCTCCATCTCCAAGGCTGGCATCGTTACC 1854

Qy    637 SerLeuAsnAlaArgThrSerValLeuAlaCysAlaAsnProSerGlySerArgTyrAsn 656
Db   1855 TCGCTGCAAGCCCGCTGCACTGTCATAGCTGCTGCCAACCCCATAGGAGGCCGCTACGAC 1914

Qy    657 AlaArgLeuSerValIleAspAsnIleGlnLeuProProThrLeuLeuSerArgPheAsp 676
Db   1915 CCTTCACTGACCTTCTCAGAGAATGTAGACCTCACAGAGCCCATCATTTCCCGCTTTGAT 1974

Qy    677 LeuIleTyrLeuMetLeuAspLysProAspGluGlnAsnAspArgArgLeuAlaArgHis 696
Db   1975 GTCCTGTGTGTGGTGAGGGACACTGTTGATCCAGTTCAGGATGAGATGCTGGCCCGCTTT 2034

Qy    697 LeuValAlaLeuHisTyrGluAsnTyrGluValSerLysGlnAsp--------------- 711
Db   2035 GTGGTTGGCAGCCACGTCAGACACCACCCCAGTAACAAGAAGGATGAAGGGTTGACTAAT 2094

Qy    712 ---------------------------------------------AlaLeuAspLeuGln 716
Db   2095 GGTGGCACCTTGGAGCCAGCCATGCCCAACACATATGGCGTGGAGCCCCTGCCTCAGGAG 2154

Qy    717 ThrLeuThrAlaTyrIleThrTyrAlaArgGlnHisValHisProThrLeuSerAspGlu 736
Db   2155 GTGCTGAAGAAGTATATCATCTATGCCAAGGAGAGGGTCCGCCCGAAGCTCAACCAGATG 2214

Qy    737 AlaAlaGluAspLeuIleAsnGlyTyrValGluMetArgGlnLysGlyAsnPheProGly 756
Db   2215 GACCAGGATAAAGTGGCCAGGATGTACAGTGACCTGAGGAAGGAGTCCATGGCAACGGGC 2274

Qy    757 SerSerLysLysValIleThrAlaThrProArgGlnLeuGluSerMetIleArgIleSer 776
Db   2275 AGC------------ATTCCCATCACGGTGCGCCACATCGAGTCCATGATCCGCATGGCC 2322

Qy    777 GluAlaLeuAlaArgMetArgPheSerGluValValGluLysValAspAlaAlaGluAla 796
Db   2323 GAGGCCCATGCCCGCATGCACCTGCGGGACTACGTGATGGAAGACGATGTCAACATGGCC 2382

Qy    797 ValArgLeuLeuAspValAlaLeuGlnGlnSerAlaThrAspHisAlaThrGlyThrIle 816
Db   2383 ATCCGAGTG------------ATGATGGAGAGCTTCATTGACACCCAGAAGTTCAGCGTC 2430

Qy    817 AspMetAspLeuIleThrThrGlyValSerAlaSerGluArgIleArgArgAlaAsn--- 835
Db   2431 ATGCGGAGTATGCGCAAGACT---TTTGCCCGGTATCTCTCCTTCCGGCGAGATAACAAT 2487

Qy    836 ------LeuLeuAlaAlaLeuArgGluLeuIleAlaAspLysIleSer 849
Db   2488 GATCTGCTGCTCTTCATACTGAAGCAGTTGGTGGCTGAGCAGGTGACA 2535
```

RESULT 5
US-11-087-227-89
; Sequence 89, Application US/11087227
; Publication No. US20050260566A1
; GENERAL INFORMATION:
; APPLICANT: Fischer, Timothy J.
; APPLICANT: Malinowski, Douglas P.
; APPLICANT: Taylor, Adriann J.
; APPLICANT: Parker, Margaret R.
; TITLE OF INVENTION: METHODS AND COMPOSITIONS FOR THE
; TITLE OF INVENTION: DETECTION OF CERVICAL DISEASE
; FILE REFERENCE: 046143/287139

```
;  CURRENT APPLICATION NUMBER: US/11/087,227
;  CURRENT FILING DATE:  2005-03-23
;  PRIOR APPLICATION NUMBER: 60/556,495
;  PRIOR FILING DATE: 2004-03-24
;  NUMBER OF SEQ ID NOS: 90
;  SOFTWARE: FastSEQ for Windows Version 4.0
; SEQ ID NO 89
;   LENGTH: 3769
;   TYPE: DNA
;   ORGANISM: Homo sapiens
US-11-087-227-89

Alignment Scores:
Pred. No.:              3.84e-55      Length:          3769
Score:                  918.50        Matches:         267
Percent Similarity:     45.43%        Conservative:    12
Best Local Similarity:  31.26%        Mismatches:      301
Query Match:            19.93%        Indels:          165
DB:                     7             Gaps:            24

US-10-768-511-8 (1-901) x US-11-087-227-89 (1-3769)

Qy        124 GlnArgSerGluLeuGlySerGlnGlyLysProLeuHisArgArgArgArgSerGlnSer 143
              :::|||      ||||||::::::                |||||:::||||||::::::
Db         48 CGCAGGCGAATCCTCGGCACTAAGCAAATATGGACCTCGCGGCGGCAGCGGAGCCGGGCG 107

Qy        144 -ArgGluProGlyHisArgSerProSerArgGluProSerAlaAspGlyArgProSerGl 163
               |||:::|||            |||   |||               ||||||
Db        108 CCGGCAGCCAGCACCTGGAGGTCCGCGACGA------------GGTGGCCGAGAAGTGCC 155

Qy        163 uSerAlaGluPro-AspAspThrLeuGlyGlyGluTyrAlaTyrValTrpGlyThrAsnV 183
                          ||| |||
Db        156 AGAAACTGTTCCTGGAC------------------------------------------- 172

Qy        183 alAsnIleProAspValLeuArgAlaIleArgArgPheLeuHisAsnTyrArgSerSerA 203
                                                  ||||||      ::::::||||||
Db        173 -----------------------------------TTCTTGGAGGAGTTTCAGAGCAGC- 196

Qy        203 laHisAspLeuAsnSerLysTyrIleGlnIleIleGluGluThrValGluArgGluGluA 223
                    |||          ||||||:::|||:::   ||||||    :::       |||    :
Db        197 -----GATGGAGAAATTAAATACTTGCAATTAGCAGAGGAACTGATTCGTCCTGAGAGAA 251

Qy        223 spThrLeuAsnIleAspMetSerAspIleTyrAspHisAspProAspLeuTyrAlaLysI 243
              ::||||||   :::          |||:::        :::      |||            |
Db        252 ACACATTGGTTGTGAGTTTTGTGGACCTGGAACAATTTAACCAGCAACTTTCCACCACCA 311

Qy        243 leValArgTyrProLeuAspIleIleProLeuLeuAspThrGluCysGlnGluValAlaT 263
              ||                       :::   |||   |||        |||:::
Db        312 TTCAAGAGGAGTTCTATAGAGTTTACCCTTACCTG---------TGTCGG---------- 352

Qy        263 hrSerLeuLeuProThrPheGluLysHisIleGluAlaArgProPhe------------- 278
                    |||   ||||||   |||       :::       |||
Db        353 -----GCCTTGAAAACATTCGTCAAAGACCGTAAAGAGATCCCTCTTGCCAAGGATTTTT 407

Qy        279 --------------AsnLeuLysAlaSerValHisMetArgGluLeuAsnProSerAspI 294
                            :::|||                 :::|||||||||      |||      |
Db        408 ATGTTGCATTCCAAGACCTGCCTACCAGACACAAGATTCGAGAGCTCACCTCATCCAGAA 467

Qy        294 leAspLysLeuValSerValLysGlyMetValIleArgCysSerSerIleIleProGluI 314
              ||       |||      :::   |||      |||:::|||             :::   ||||||:
Db        468 TTGGTTTGCTCACTCGCATCAGTGGGCAGGTGGTGCGGACTCACCCAGTTCACCCAGAGC 527

Qy        314 leLysGlyAlaPhePheLysCysLeuValCysGlyHisSerProProLeuValThrValV 334
              ::                |||      ||||||      |||                      ||||||:
Db        528 TTGTGAGCGGAACTTTTCTGTGCTTGGACTGT-------------------CAGACAGTGA 569

Qy        334 alLysGly---------------ArgValGluGluProThrArgCysGluLysProGluC 349
              :::::                   :::           :::|||      |||         |||      |
Db        570 TCAGGGATGTAGAACAGCAGTTCAAATACACACAGCCAAACATCTGCCGAAATCCAGTTT 629

Qy        349 ysAlaAlaArgAsnAlaMetSerLeuIleHisAsnArgCysThrPheAlaAsnLysGlnI 369
              |||||   |||               |||      |||:::      |||    :::   |||
Db        630 GTGCCAACAGGAGGAGATTCTTACTGGATACAAATAAATCAAGATTTGTTGATTTTCAAA 689

Qy        369 leValArgLeuGlnGluThrProAspAlaIleProGluGlyGluThrProHisThrValS 389
                |||||:::||||||||          :::|||   |||         |||      ::::::
Db        690 AGGTTCGTATTCAAGAGACCCAAGCTGAGCTTCCTCGAGGGAGTATCCCCGCAGTTTAG 749

Qy        389 erMetCysLeuTyrAsnThrMetValAspAlaValLysProGlyAspArgIleGluValT 409
                :::      |||               |||::::::   :::      ||||||:::    :::    |
Db        750 AAGTAATTTTAAGGGCTGAAGCTGTGGAATCAGCTCAAGCTGGTGACAAGTGTGACTTTA 809

Qy        409 hrGlyVal---------------------------PheLysAlaMetAlaValArgValG 420
              |||||                                                       ||
Db        810 CAGGGACACTGATTGTTGTGCCTGACGTCTCCAAGCTTAGCACACCAGGAGCACGTGCAG 869

Qy        420 lyProAsnGlnArgThr---------------------------------------LeuA 427
                    |||   |||                                          ||||
Db        870 AAACTAATTCCCGTGTCAGTGGTGTTGATGGATATGAGACAGAAGGCATTCGAGGACTCC 929

Qy        427 rgAlaLeu-----------------TyrLys---ThrTyrIleAspCysValHisValL 440
              ||||||||                 |||:::       ::::::   |||
Db        930 GGGCCCTTGGTGTTAGGGACCTTTCTTATAGGCTGGTCTTTCTTGCCTGCTGT------- 982

Qy        440 ysLysSerAspArgGlyArgLeuGlnThrGluAspProMetGluMetAspLysGluAsnA 460
                                                               :::   |||
Db        983 ------------------------------------------GTTGCGCCAACCAACC 998

Qy        460 spMetTyrAlaGlyTyrHisGluSerAspThrSerGluAlaAlaAsnGluAlaLysIleG 480
                 :::   |||               |||      :::      |||            |||   |
Db        999 CAAGGTTTGGGGGGAAAGAGCTCAGAGATGAGGAACAGACAGCTGAGAGCATTAAGAACC 1058

Qy        480 ln--------------------LysLeuLysGluLeuSerLysLeuProGlyIleTyrA 493
              ||                    |||:::   |||:::|||:::            :::|||
Db       1059 AAATGACTGTGAAAGAATGGGAGAAAGTGTTTGAGATGAGTCAAGATAAAAATCTATACC 1118

Qy        493 spArgLeuSerArgSerLeuAlaProSerIleTrpGluLeuGluAspIleLysLysGlyL 513
                 |||      ||||||      |||:::|||               ::::::::|||:::|||:
Db       1119 ACAATCTTTGTACCAGCCTGTTCCCTACTATACATGGCAATGATGAAGTAAAACGGGGTG 1178

Qy        513 euLeuCysGlnLeuPheGlyGlyLysAlaLysLysIleProSerGlyAlaSerPheArgG 533
              :::|||      |||||||||||             |||      |||   |||   |||
Db       1179 TCCTGCTGATGCTCTTTGGTGGCGTTCCAAAGACAACAGGAGAAGGGACCTCTCTTCGAG 1238

Qy        533 lyAspIleAsnValLeuLeuValGlyAspProGlyThrSerLysSerGlnLeuLeuGlnT 553
              ||||||||||||   :::|||||||||||   |||:::||||||||   |||::::
Db       1239 GGACATAAATGTTTGCATTGTTGGTGACCCAAGTACAGCTAAGAGCCAATTTCTCAAGC 1298

Qy        553 yrValHisLysIleAlaProArgGlyIleTyrThrSerGlyArgGlySerSerAlaValG 573
              :::|||   :::   :::|||||   :::|||||||||||:::   ||||||||      |
Db       1299 ACGTGGAGGAGTTCAGCCCCAGAGCTGTCTACACCAGTGGTAAAGCGTCCAGTGCTGCTG 1358

Qy        573 lyLeuThrAlaTyrValThrLysAspProGluThrArgGluThrValLeuGluSerGlyA 593
              ||||||||||   |||   :::|||   |||:::   |||   |||:::|||:::||||
Db       1359 GCTTAACAGCAGCTGTTGTGAGAGATGAAGAATCTCATGAGTTTGTCATTGAGGCTGGAG 1418

Qy        593 laLeuValLeuSerAspArgGlyIleCysCysIleAspGluPheAspLysMetSerAspA 613
              |||||:::|||:::|||   |||:::||||||||||||||||||||||||||||||||
Db       1419 CTTTGATGTTGGCTGATAATGGTGTGTGTTGTATTGATGAATTTGATAAGATGGACGTGC 1478

Qy        613 snAlaArgSerMetLeuHisGluValMetGluGlnGlnThrValSerValAlaLysGlyG 633
                 :::      :::||||||   ||||||||||||||:::|||:::      |||   |
Db       1479 GGGATCAAGTTGCTATTCATGAAGCTATGGAACAGCAGACCATATCCATCACTAAAGCAG 1538

Qy        633 lyIleIleAlaSerLeuAsnAlaArgThrSerValLeuAlaCysAlaAsnProSerGlyS 653
              |||:::   |||:::|||||||||||||||||||:::||||||   |||||||||
Db       1539 GAGTGAAGGCTACTCTGAACGCCCGGACGTCCATTTTGGCAGCAGCAAACCCAATCAGTG 1598

Qy        653 erArgTyrAsnAlaArgLeuSerValIleAspAsnIleGlnLeuProProThrLeuLeuS 673
                 |||:::         |||:::         ||||||   |||         ::::::|
Db       1599 GACACTATGACAGATCAAAATCATTGAAACAGAATATAAATTTGTCAGCTCCCATCATGT 1658

Qy        673 erArgPheAspLeuIleTyrLeuMetLeuAspLysProAspGluGlnAsnAspArgArgL 693
              |||||||||||||   ::::::::::::|||:::      :::|||         |||        :
Db       1659 CCCGATTCGATCTCTTCTTTATCCTTGTGGATGAATGTAATGAGGTTACAGATTATGCCA 1718
```

```
Qy        693 euAlaArgHisLeuValAlaLeuHisTyrGluAsnTyrGluValSerLysGlnAspAlaL 713
              ::|||||   :::|||  ||||||              |||  |||
Db       1719 TTGCCAGGCGCATAGTAGATTTGCAT---TCAAGAATTGAGGAATCAATTGATCGTGTCT 1775

Qy        713 euAspLeuGlnThrLeuThrAlaTyrIleThrTyrAlaArgGlnHisValHisProThrL 733
                 |||      :::      |||:::   :::||||||||          |||    :
Db       1776 ATTCCCTCGATGATATCAGAAGATATCTTCTCTTTGCAAGACAG---TTTAAACCCAAGA 1832

Qy        733 euSerAspGluAlaAlaGluAspLeuIleAsnGlyTyrValGluMetArgGlnLysGlyA 753
              ::|||   |||:::   :::   ::::::       |||        :::||||||:::
Db       1833 TTTCCAAAGAGTCAGAGGACTTCATTGTGGAGCAATATAAACATCTCCGCCAGAGAGATG 1892

Qy        753 snPheProGlySerSerLysLysValIleThrAlaThrProArgGlnLeuGluSerMetI 773
                    |||   :::|||               |||   |||||||||||||||||
Db       1893 GT---TCTGGAGTGACCAAGTCTTCATGGAGGATTACAGTGCGACAGCTTGAGAGCATGA 1949

Qy        773 leArgIleSerGluAlaLeuAlaArgMetArgPheSerGluValValGluLysValAspA 793
              |||||:::||||||||:::||||||||             :::    |||:::
Db       1950 TTCGTCTCTCTGAAGCTATGGCTCGGATGCACTGCTGTGATGAGGTCCAACCTAAACATG 2009

Qy        793 laAlaGluAlaValArgLeuLeu-----------------------------AspValA 803
                    ||||||  |||||||||                            ||||||
Db       2010 TGAAGGAAGCTTTCCGGTTACTGAATAAATCAATCATCCGTGTGGAAACACCTGATGTCA 2069

Qy        803 laLeuGlnGlnSer---------------AlaThrAspHisAlaThrGlyThrIle---- 816
                |||  |||                         |||        |||   |||
Db       2070 ATCTAGATCAAGAGGAAGAGATCCAGATGGAGGTAGATGAGGGTGCTGGTGGCATCAATG 2129

Qy        817 --------AspMetAspLeuIleThrThrGlyValSerAla------------------ 827
                      |||                   |||::::::
Db       2130 GTCATGCTGACAGCCCTGCTCCTGTGAACGGGATCAATGGCTACAATGAAGACATAAATC 2189

Qy        828 -------------------------------------SerGluArgIleArgArgAlaA 835
                                                   ||||||     |||    :::|
Db       2190 AAGAGTCTGCTCCCAAAGCCTCCTTAAGGCTGGGCTTCTCTGAGTACTGCCGAATCTCTA 2249

Qy        835 snLeuLeuAlaAla---LeuArgGluLeuIleAlaAspLysIleSerProGlySerSerS 854
              |||||:::         ||||||::::::     ::::::                  |
Db       2250 ACCTTATTGTGCTTCACCTCAGAAAGGTGGAAGAAGAAGAG------------GACGAGT 2297

Qy        854 erGlyLeuLysThrSerGlnLeu------------LeuGluAspIleArgSerGlnSerS 870
              ||   ||||||   |||:::|||            |||::::::|||   |||:::
Db       2298 CAGCATTAAAGAGGAGCGAGCTTGTTAACTGGTACTTGAAGGAAATCGAATCAGAGATAG 2357

Qy        870 erValAspValSerLeuGlnAspIleLysAsnAlaLeu 882
                    :::      |||    :::    |||        :::
Db       2358 ACTCTGAAGAAGAACTTATAAATAAAAAAAGAATCATA 2395
```

RESULT 6

US-10-507-275-2

```
; Sequence 2, Application US/10507275
; Publication No. US20050250166A1
; GENERAL INFORMATION:
;  APPLICANT: Masai, Hisao
;  APPLICANT:  Tamai, Katsuyuki
;  APPLICANT:  Medical and Biological Laboratories Co., Ltd.
;  APPLICANT:  Japan Science and Technology Agency
;  APPLICANT:  Ginkgo Biomedical Research Institute Co., Ltd.
;  TITLE OF INVENTION: Cdc7-ASK Kinase Complex, Substrates of the Kinase Complex,
;  TITLE OF INVENTION:  Specific Antibodies to the Substrates, and Screening Methods
;  TITLE OF INVENTION:  Using the Same to Screen for Compounds Comprising Cdc7-ASK
;  TITLE OF INVENTION:  Kinase Inhitiory Ability
;  FILE REFERENCE: 082368-001100US
;  CURRENT APPLICATION NUMBER: US/10/507,275
;  CURRENT FILING DATE:  2004-09-09
;  PRIOR APPLICATION NUMBER: JP 2002-067702
;  PRIOR FILING DATE: 2002-03-12
;  PRIOR APPLICATION NUMBER: WO PCT/JP03/02918
;  PRIOR FILING DATE: 2003-03-12
;  NUMBER OF SEQ ID NOS: 21
;  SOFTWARE: PatentIn Ver. 2.1
; SEQ ID NO 2
;   LENGTH: 3379
;   TYPE: DNA
;   ORGANISM: Homo sapiens
;   FEATURE:
;   NAME/KEY: CDS
;   LOCATION: (31)..(2709)
US-10-507-275-2

Alignment Scores:
Pred. No.:                 8.92e-51        Length:          3379
Score:                     856.00          Matches:         244
Percent Similarity:        45.26%          Conservative:    143
Best Local Similarity:     28.54%          Mismatches:      322
Query Match:               18.57%          Indels:          147
DB:                        6               Gaps:            19

US-10-768-511-8 (1-901) x US-10-507-275-2 (1-3379)

Qy         31 ThrSerProSerPheAspAsnAlaAlaSerProValAlaGlyArgArgAlaValArgGln 50
                :::|||   ||||||      |||:::||||||   |||   |||||
Db         13 TCATCGGAATCCTTCACCATGGCATCCAGCCCG---GCCCAGCGTCGGCGAGGCAATGAT 69

Qy         51 ThrProThrSerAlaValArgArgArgGlyArgGluThrAspSerAlaArgArgArgArg 70
                    ||||||:::          |||      |||   ||||||:::
Db         70 CCTCTCACCTCCAGCCCTGGCCGAAGCTCCCGGCGTACTGATGCCCTC------ACCTCC 123

Qy         71 SerArgSerArgSerLeuGlyAsnSerValTyrSerSerProTyrAspAlaGlyThrPro 90
                |||      |||   |||                      |||::::::      :::
Db        124 AGCCCTGGCCGTGACCTT------------------CCACCATTTGAGGATGAGTCCGAG 165

Qy         91 GlyThrProGlyThrProValAlaThrProValTyrAlaThrProValGlyThrProMet 110
              |||      ||||||                                |||   |||
Db        166 GGGCTCCTAGGCACA--------------------GAGGGGCCCCT-GGAGGAAGAAGA 203

Qy        111 GlyThrProSerPheHis---------------ArgGlyThrProGlnTyrLysGlnArg 125
              |||            |||                  :::|||   |||::::::   :::
Db        204 GGATGGAGAGGAGCTCATTGGAGATGGCATGGAAAGGGACTACCGCGCCATCCCAGAGCT 263

Qy        126 SerGluLeu--------------------------------------------------- 128
                    |||
Db        264 GGACGCCTATGAGGCCGAGGGACTGGCTCTGGATGATGAGGACGTAGAGGAGCTGACGGC 323

Qy        129 ------------GlySerGlnGlyLysProLeuHis------------------------ 136
                          ||||||:::                |||
Db        324 CAGTCGAAGGGAGGCAGCAGACGGGCCATGCGGCACGGTGACCGGGAGCTGGCCGGGGCT 383

Qy        136 ------------------------------------------------------------ 136

Db        384 GGGCGCATGCGCCGTGGGCTCCTGTATGACAGCGATGAGGAGGACGAGGAGCGCCCTGCC 443

Qy        137 ArgArgArgArgSerGlnSerArgGluProGlyHisArgSerProSerArgGluProSer 156
              |||:::||||||            ||||||         |||    |||        :::
Db        444 CGCAAGCGCCGCCAGTGGAGCCGGCACGGAGGACGGCGAGGAGGACGAGCAGATGATTGA 503

Qy        157 AlaAspGlyArgProSerGluSerAlaGluProAspAspThrLeuGlyGlyGluTyrAla 176
                       |||      |||         |||                ||||||:::
Db        504 GAGCATCGAGAACCTGGAGGATCTCAAAGGCCACTCTGTGCGCGAGTGGGTGAGCATGGC 563

Qy        177 TyrValTrpGlyThrAsnValAsnIleProAspValLeuArgAlaIleArgArgPheLeu 196
                       |||                  |||        |||:::   :::
Db        564 ---------GGGCCCCCGGCTGGAGATCCACCACCGCTTCAAGAACTTCCT------GCG 608

Qy        197 HisAsnTyrArg-SerSerAlaHisAspLeuAsnSerLysTyrIleGlnIleIleGluGl 216
              |||:::   |||  |||         |||::::::             :::   :::   |||  ::
Db        609 CACTCACGTCGACAGCCACGGCCACAACGTC---------TTCAAGGAGCGCATCAGCGA 659

Qy        216 uThrValGluArgGluGluAspThrLeuAsnIleAspMetSerAspIleTyrAspHisAs 236
              :       :::            ::::::|||   ::::::        |||:::      ::
Db        660 CATGTGCAAAGAGAACCGTGAGAGCCTGGTGGTGAACTATGAGGACTTGGCAGCCAGGGA 719

Qy        236 pProAspLeuTyrAlaLysIleValArgTyrProLeuAspIleIleProLeuLeuAspTh 256
              :        |||            :::         |||   :::::::::   :::   |||
Db        720 GCACGTGCTGGCCTACTTCCTGCCTGAGGCACCGGCGGAGCTGCTGCAGATCTTTGATGA 779
```

```
Qy        256 rGluCysGlnGluValAlaThrSerLeuLeuProThrPheGluLys---------HisIl 273
Db        780 GGCTGCCCTGGAGGTGGTACTGGCCATGTACCCCAAGTACGACCGCATCACCAACCACAT 839

Qy        273 eGluAlaArgProPheAsnLeuLysAlaSerValHisMetArgGluLeuAsnProSerAs 293
Db        840 CCATGTCCGCATCTCCCACCTGCCTCTGGTGGAGGAGCTGCGCTCGCTGAGGCAGCTGCA 899

Qy        293 pIleAspLysLeuValSerValLysGlyMetValIleArgCysSerSerIleIleProGl 313
Db        900 TCTGAACCAGCTGATCCGCACCAGTGGGGTGGTGACCAGCTGCACTGGCGTCCTGCCCCA 959

Qy        313 uIleLysGlyAlaPhePheLysCysLeuValCysGlyHisSerProProLeuValThrVa 333
Db        960 GCTCAGCATGGTCAAGTACAACTGCAACAAGTGCAATTTCGTCCTGGGTCCTTTCTGCCA 1019

Qy        333 lValLysGlyArgValGluGluProThrArgCysGluLysProGluCysAlaAlaArgAs 353
Db       1020 GTCCCAGAACCAGGAGGTGAAACCAGGCTCCTGT------CCTGAGTGCCAGTCGGCCGG 1073

Qy        353 nAlaMetSerLeuIleHisAsnArgCysThrPheAlaAsnLysGlnIleValArgLeuGl 373
Db       1074 CCCCTTTGAGGTCAACATGGAGGAGACCATCTATCAGAACTACCAGCGTATCCGAATCCA 1133

Qy        373 nGluThrProAspAlaIleProGluGlyGluThrProHisThrValSerMetCysLeuTy 393
Db       1134 GGAGAGTCCAGGCAAAGTGGCGGCTCGGCGGCTGCCCCGCTCCAAGGACGCCATTCTCCT 1193

Qy        393 rAsnThrMetValAspAlaValLysProGlyAspArgIleGluValThrGlyValPheLy 413
Db       1194 CGCAGATCTGGTGGACAGCTGCAACGCAGGAGACGAGATAGAGCTGACTGGCATCTATCA 1253

Qy        413 sAlaMetAlaValArgValGlyProAsnGlnArgThrLeuArgAla------LeuTyrLy 431
Db       1254 CAAC---------AACTATGATGGCTCCCTCAACACTGCCAATGGCTTCCCTGTCTTTGC 1304

Qy        431 sThrTyrIleAspCysValHisValLysLysSerAspArg------------GlyArgLe 447
Db       1305 CACTGTCATCCTAGCCAACCACGTGGCCAAGAAGGACAACAAGGTTGCTGTAGGGGAACT 1364

Qy        447 uGlnThrGluAspProMetGluMetAspLysGluAsnAspMetTyrAlaGlyTyrHisGl 467
Db       1365 GACCGATGAAGAT----------------------------------------------- 1377

Qy        467 uSerAspThrSerGluAlaAlaAsnGluAlaLysIleGlnLysLeuLysGluLeuSerLy 487
Db       1378 ----------------------------------GTGAAGATGATCACTAGCCTCTCCAA 1403

Qy        487 sLeuProGlyIleTyrAspArgLeuSerArgSerLeuAlaProSerIleTrpGluLeuGl 507
Db       1404 GGATCAGCAGATCGGAGAGAAGATCTTTGCCAGCATTGCTCCTTCCATCTATGGTCATGA 1463

Qy        507 uAspIleLysLysGlyLeuLeuCysGlnLeuPheGlyGlyLysAlaLysLysIleProSe 527
Db       1464 AGACATCAAGAGAGGCCCTGCTCTGGCCCTGTTCGGAGGGGAGCCCAAAAACCCAGGTGG 1523

Qy        527 rGlyAlaSerPheArgGlyAspIleAsnValLeuLeuValGlyAspProGlyThrSerLy 547
Db       1524 CAAGCACAAGGTACGTGGTGATATCAACGTGCTCTTGTGCGGAGACCCTGGCACAGCGAA 1583

Qy        547 sSerGlnLeuLeuGlnTyrValHisLysIleAlaProArgGlyIleTyrThrSerGlyAr 567
Db       1584 GTCGCAGTTTCTCAAGTATATTGAGAAAGTGTCCAGCCGAGCCATCTTCACCACTGGCCA 1643

Qy        567 gGlySerSerAlaValGlyLeuThrAlaTyrValThrLysAspProGluThrArgGluTh 587
Db       1644 GGGGCGTCGGCTGTGGCCGTCACGGCGTATGTCCAGCGGCACCCTGTCAGCAGGGAGTG 1703

Qy        587 rValLeuGluSerGlyAlaLeuValLeuSerAspArgGlyIleCysCysIleAspGluPh 607
Db       1704 GACCTTGGAGGCTGGGGCCCTGGTTCTGGCTGACCGAGGAGTGTGTCTCATTGATGAATT 1763

Qy        607 eAspLysMetSerAspAsnAlaArgSerMetLeuHisGluValMetGluGlnGlnThrVa 627
Db       1764 TGACAAGATGAATGACCAGGACAGAACCAGCATCCATGAGGCCATGGAGCAACAGAGCAT 1823

Qy        627 lSerValAlaLysGlyGlyIleIleAlaSerLeuAsnAlaArgThrSerValLeuAlaCy 647
Db       1824 CTCCATCTCGAAGGCTGGCATCGTCACCTCCCTGCAGGCTCGCTGCACGGTCATTGCTGC 1883

Qy        647 sAlaAsnProSerGlySerArgTyrAsnAlaArgLeuSerValIleAspAsnIleGlnLe 667
Db       1884 CGCCAACCCCATAGGAGGGCGCTACGACCCCTCGCTGACTTTCTCTGAGAACGTGGACCT 1943

Qy        667 uProProThrLeuLeuSerArgPheAspLeuIleTyrLeuMetLeuAspLysProAspGl 687
Db       1944 CACAGAGCCCATCATCTCACGCTTTGACATCCTGTGTGTGGTGAGGGACACCGTGGACCC 2003

Qy        687 uGlnAsnAspArgArgLeuAlaArgHisLeuValAlaLeuHisTyrGluAsnTyrGluVa 707
Db       2004 AGTCCAGGACGAGATGCTGGCCCGCTTCGTGGTGGGCAGCCACGTCAGACACCACCCCAG 2063

Qy        707 lSerLysGln-------------------------------------------------- 710
Db       2064 CAACAAGGAGGAGGAGGGGCTGGCCAATGGCAGCGCTGCTGAGCCCGCCATGCCCAACAC 2123

Qy        711 ----------AspAlaLeuAspLeuGlnThrLeuThrAlaTyrIleThrTyrAlaArgGl 727
Db       2124 GTATGGCGTGGAGCCCCTGCCCCAGGAGGTCCTGAAGAAGTACATCATCTACGCCAAGGA 2183

Qy        727 nHisValHisProThrLeuSerAspGluAlaAlaGluAspLeuIleAsnGlyTyrValGl 747
Db       2184 GAGGGTCCACCCGAAGCTCAACCAGATGGACCAGGACAAGGTGGCCAAGATGTACAGTGA 2243

Qy        747 uMetArgGlnLysGlyAsnPheProGlySerSerLysLysValIleThrAlaThrProAr 767
Db       2244 CCTGAGGAAAGAATCTATGGCGACAGGCAGC------------ATCCCCATTACGGTGCG 2291

Qy        767 gGlnLeuGluSerMet-IleArgIleSerGluAlaLeuAlaArgMetArgPheSerGluV 787
Db       2292 GCACATCGAGTCCATGAGTC---ATGGCGGAGGCCCACGCGCGCATCCATCTGCGGGACT 2348

Qy        787 alValGluLysValAspAlaAlaGluAlaValArgLeuLeu 800
Db       2349 ATGTGATCGAAGACGACGTCAACATGGCCATCCGCGTGATG 2389
```

RESULT 7
US-11-108-172-158
; Sequence 158, Application US/11108172
; Publication No. US20050260177A1
; GENERAL INFORMATION:
; APPLICANT: Xu, Jiangchun
; APPLICANT: Lodes, Michael J.
; APPLICANT: Secrist, Heather
; APPLICANT: Benson, Darin R.
; APPLICANT: Meagher, Madeleine Joy
; APPLICANT: Stolk, John A.
; APPLICANT: Wang, Tongtong
; APPLICANT: Jiang, Yuqiu
; APPLICANT: Smith, Carole L.
; APPLICANT: King, Gordon E.
; APPLICANT: Wang, Aijun
; APPLICANT: Clapper, Jonathan D.
; APPLICANT: Skeiky, Yasir A. W.
; APPLICANT: Fanger, Gary R.
; APPLICANT: Vedvick Thomas S.
; APPLICANT: Carter, Darrick
; TITLE OF INVENTION: COMPOUNDS FOR IMMUNOTHERAPY AND DIAGNOSIS
; TITLE OF INVENTION: OF COLON CANCER AND METHODS FOR THEIR USE
; FILE REFERENCE: 210121.471C15
; CURRENT APPLICATION NUMBER: US/11/108,172
; CURRENT FILING DATE: 2005-04-15
; PRIOR APPLICATION NUMBER: US 10/025,380
; PRIOR FILING DATE: 2001-12-19
; PRIOR APPLICATION NUMBER: US 09/922,217
; PRIOR FILING DATE: 2001-08-03

```
;  PRIOR APPLICATION NUMBER: US 09/833,263
;  PRIOR FILING DATE: 2001-04-10
;  PRIOR APPLICATION NUMBER: US 09/649,811
;  PRIOR FILING DATE: 2000-08-28
;  PRIOR APPLICATION NUMBER: US 09/609,448
;  PRIOR FILING DATE: 2000-06-29
;  PRIOR APPLICATION NUMBER: US 09/575,251
;  PRIOR FILING DATE: 2000-05-19
;  PRIOR APPLICATION NUMBER: US 09/519,444
;  PRIOR FILING DATE: 2000-03-06
;  PRIOR APPLICATION NUMBER: US 09/504,629
;  PRIOR FILING DATE: 2000-02-15
;  PRIOR APPLICATION NUMBER: US 09/480,321
;  PRIOR FILING DATE: 2000-01-10
;  PRIOR APPLICATION NUMBER: US 09/476,296
;  PRIOR FILING DATE: 1999-12-30
;  Remaining Prior Application data removed - See File Wrapper or PALM.
;  NUMBER OF SEQ ID NOS: 1130
;  SOFTWARE: FastSEQ for Windows Version 4.0
; SEQ ID NO 158
;   LENGTH: 507
;   TYPE: DNA
;   ORGANISM: Homo sapiens
;   FEATURE:
;   NAME/KEY: misc_feature
;   LOCATION: 27, 428, 448, 449, 456, 462, 490, 492, 497, 498, 502, 503
;   OTHER INFORMATION: n = A,T,C or G
US-11-108-172-158

Alignment Scores:
Pred. No.:              3e-22        Length:         507
Score:                  438.50       Matches:        90
Percent Similarity:     64.29%       Conservative:   27
Best Local Similarity:  49.45%       Mismatches:     48
Query Match:            9.51%        Indels:         17
DB:                     7            Gaps:           2

US-10-768-511-8 (1-901) x US-11-108-172-158 (1-507)

Qy       410 GlyValPheLysAlaMetAlaValArgValGlyProAsnGlnArgThrLeuArgAlaLeu 429
Db         1 GGCACGAGTCGAGCTGTGCCTATTCGNGTCAATCCAAGAGTGAGTAATGTGAAGTCTGTC 60

Qy       430 TyrLysThrTyrIleAspCysValHisValLysLysSerAspArgGlyArgLeuGlnThr 449
Db        61 TACAAAACCCACATTGATGTCATTCATTATCGGAAAACGGATGCAAAACGTCTGCAT--- 117

Qy       450 GluAspProMetGluMetAspLysGluAsnAspMetTyrAlaGlyTyrHisGluSerAsp 469
Db       118 ------------------------------------------GGCCTTGATGAAGAAGCA 135

Qy       470 ThrSerGluAlaAlaAsnGluAlaLysIleGlnLysLeuLysGluLeuSerLysLeuPro 489
Db       136 GAACAGAAACTTTTTTCAGAGAAACGTGTGGAATTGCTTAAGGAACTTTCCAGGAAACCA 195

Qy       490 GlyIleTyrAspArgLeuSerArgSerLeuAlaProSerIleTrpGluLeuGluAspIle 509
Db       196 GACATTTATGAGAGGCTTGCTTCAGCCTTGGCTCCAAGCATTTATGAACATGAAGATATA 255

Qy       510 LysLysGlyLeuLeuCysGlnLeuPheGlyGlyLysAlaLysLysIle------ProSer 527
Db       256 AAGAAGGGAATTTTGCTTCAGCTCTTTGGCGGACAAGGAAGGATTTAGTCACACTGGA 315

Qy       528 GlyAlaSerPheArgGlyAspIleAsnValLeuLeuValGlyAspProGlyThrSerLys 547
Db       316 AGGGGCAAATTTCGGGCTGAGATCAACATCTTGCTGTGTGGCGACCCTGGTACCAGCAAG 375

Qy       548 SerGlnLeuLeuGlnTyrValHisLysIleAlaProArgGlyIleTyrThrSerGlyArg 567
Db       376 TCCCAGCTGCTGCAGTACGTGTACAACCTCGTCCCCAGGGGCCAGTACACGTNTGGGAAG 435

Qy       568 GlySerSerAlaValGlyLeuThrAlaTyrValThrLysAspProGluThrArgGluThr 587
Db       436 GGCTCCAGTGCANNTGGCCTNACTGCNTACGTAATGAAAGACCCTGAGACAAGGNANCTG 495

Qy       588 ValLeu 589
Db       496 GNNCTG 501

RESULT 8
US-10-750-185-55669
; Sequence 55669, Application US/10750185
; Publication No. US20050260603A1
; GENERAL INFORMATION:
;  APPLICANT: MMI GENOMICS, INC.
;  APPLICANT:  DeNISE, Sue K.
;  APPLICANT:  KERR, Richard
;  APPLICANT:  ROSENFELD, David
;  APPLICANT:  HOLM, Tom
;  APPLICANT:  BATES, Stephen
;  APPLICANT:  FANTIN, Dennis
;  TITLE OF INVENTION: COMPOSITIONS FOR INFERRING BOVINE TRAITS
;  FILE REFERENCE: MMI1100-2
;  CURRENT APPLICATION NUMBER: US/10/750,185
;  CURRENT FILING DATE:  2003-12-31
;  PRIOR APPLICATION NUMBER: US 60/437,482
;  PRIOR FILING DATE: 2002-12-31
;  NUMBER OF SEQ ID NOS: 64922
;  SOFTWARE: PatentIN version 3.1
; SEQ ID NO 55669
;   LENGTH: 1312
;   TYPE: DNA
;   ORGANISM: Bovine  19866880618533
US-10-750-185-55669

Alignment Scores:
Pred. No.:              0.00353      Length:         1312
Score:                  176.50       Matches:        37
Percent Similarity:     66.67%       Conservative:   11
Best Local Similarity:  51.39%       Mismatches:     17
Query Match:            3.83%        Indels:         7
DB:                     6            Gaps:           1

US-10-768-511-8 (1-901) x US-10-750-185-55669 (1-1312)

Qy       562 IleTyrThrSerGlyArgGlySerSerAlaValGlyLeuThrAlaTyrValThrLysAsp 581
Db       840 GTGTACACGTCTGGGAAAGGCAGCAGCGCCGCCGGGCTGACAGCCTCGGTGATGAGGGAC 899

Qy       582 ProGluThrArgGluThrValLeuGluSerGlyAlaLeuValLeuSerAspArgGlyIle 601
Db       900 CCCTCGTCCCGGAACTTCATCATGGAGGGCGGGGCCATGGTCCTGGCCGACGGAGGTGTC 959

Qy       602 CysCysIleAspGluPheAspLysMetSerAspAsnAlaArgSerMetLeuHisGluVal 621
Db       960 GTCTGTATTGATGAGTTTGACAAGGTGAGCCTGGCAGGGCGAGGG--------------- 1004

Qy       622 MetGluGlnGlnThrValSerValAlaLysGlyGly 633
Db      1005 ------GGTGGCTCAGTGCTGGTGGCACCTGGTGGC 1034

RESULT 9
US-10-750-185-45476
; Sequence 45476, Application US/10750185
; Publication No. US20050260603A1
; GENERAL INFORMATION:
;  APPLICANT: MMI GENOMICS, INC.
;  APPLICANT:  DeNISE, Sue K.
;  APPLICANT:  KERR, Richard
;  APPLICANT:  ROSENFELD, David
;  APPLICANT:  HOLM, Tom
;  APPLICANT:  BATES, Stephen
;  APPLICANT:  FANTIN, Dennis
;  TITLE OF INVENTION: COMPOSITIONS FOR INFERRING BOVINE TRAITS
;  FILE REFERENCE: MMI1100-2
;  CURRENT APPLICATION NUMBER: US/10/750,185
```

```
;  CURRENT FILING DATE:  2003-12-31
;  PRIOR APPLICATION NUMBER: US 60/437,482
;  PRIOR FILING DATE: 2002-12-31
;  NUMBER OF SEQ ID NOS: 64922
;  SOFTWARE: PatentIN version 3.1
; SEQ ID NO 45476
;   LENGTH: 1851
;   TYPE: DNA
;   ORGANISM: Bovine  19866880565574
US-10-750-185-45476

Alignment Scores:
Pred. No.:              0.00817      Length:        1851
Score:                  174.00       Matches:       34
Percent Similarity:     74.60%       Conservative:  13
Best Local Similarity:  53.97%       Mismatches:    14
Query Match:            3.78%        Indels:        2
DB:                     6            Gaps:          1

US-10-768-511-8 (1-901) x US-10-750-185-45476 (1-1851)

Qy        571 AlaValGlyLeuThrAlaTyrValThrLysAspProGluThrArgGluThrValLeuGlu 590
              ::::::|||||||||          ||||||          :::   |||    ||||||
Db       1366 TCCTTAGGTCTGACGGTAACTGCTGTAAAAGAC------TCAGGAGAATGGAATTTGGAG 1419

Qy        591 SerGlyAlaLeuValLeuSerAspArgGlyIleCysCysIleAspGluPheAspLysMet 610
              :::||||||||||||||:::|||   |||:::||||||||||||||||||:::    :::
Db       1420 GCGGGAGCATTAGTCCTCGCAGATGCTGGCCTCTGCTGTATCGATGAATTTAATAGCCTC 1479

Qy        611 SerAspAsnAlaArgSerMetLeuHisGluValMetGluGlnGlnThrValSerValAla 630
                 ::::::   |||:::   :::||||||   ||||||||||||||:::|||||||||
Db       1480 AAAGAGCATGACAGAACTAGTATCCACGAAGCAATGGAGCAGCAAACCATCAGTGTTGCT 1539

Qy        631 LysGlyGly 633
              |||   |||
Db       1540 AAGGCTGGG 1548

RESULT 10
US-11-004-057-3
; Sequence 3, Application US/11004057
; Publication No. US20050244846A1
; GENERAL INFORMATION:
;  APPLICANT: Johnson, Gary L.
;  TITLE OF INVENTION: MEKK1 PROTEINS AND FRAGMENTS THEREOF FOR USE IN REGULATING
;  TITLE OF INVENTION:  APOPTOSIS
;  FILE REFERENCE: CPI-042CPPC
;  CURRENT APPLICATION NUMBER: US/11/004,057
;  CURRENT FILING DATE:  2004-12-02
;  PRIOR APPLICATION NUMBER: US/09/403,075
;  PRIOR FILING DATE: 2000-05-10
;  NUMBER OF SEQ ID NOS: 21
;  SOFTWARE: PatentIn Ver. 2.0
; SEQ ID NO 3
;   LENGTH: 5253
;   TYPE: DNA
;   ORGANISM: Mus musculus
;   FEATURE:
;   NAME/KEY: CDS
;   LOCATION: (15)..(4496)
US-11-004-057-3

Alignment Scores:
Pred. No.:              1.52         Length:        5253
Score:                  150.00       Matches:       224
Percent Similarity:     32.62%       Conservative:  125
Best Local Similarity:  20.93%       Mismatches:    364
Query Match:            3.25%        Indels:        358
DB:                     9            Gaps:          56

US-10-768-511-8 (1-901) x US-11-004-057-3 (1-5253)

Qy         11 AlaValSerSerProTyrProSerGlnSerGluGlyValSerThrProLeuProGlnVal 30
              |||   ||||||   :::|||                |||   :::         |||:::
Db         39 GCCTCGTCGTCGGGATTCCCG------------GGCGCCGCGGCGGCGAGTCCCGAGGCG 86

Qy         31 ---------------ThrSerProSerPheAspAsnAlaAlaSerProValAlaGlyArg 45
                                            :::       :::   :::|||   ||||||
Db         87 GGCGGCGGCGGCGGAGGAGGAGGAGCTCTCCAGGGAAGCGGCGCGCCCGCAGCGGGCGCG 146

Qy         46 ArgAlaValArgGlnThrProThrSerAlaValArgArgArgGlyArgGluThrAspSer 65
                 :::   :::   |||   ||||||                  |||||||||
Db        147 GCGGGGCTGCTGCGGGAGCCTGGCAGCGCG-----------GGCCGCGAGCGCGCGGAC 194

Qy         66 AlaArgArgArgArgSerArgSerArgSerLeuGlyAsnSerValTyrSerSerProTyr 85
                 |||||||||      |||                        |||   |||
Db        195 TGGCGGCGGCGGCACGTGCGC-----------------AAAGTGCGGAGTGTGGAGCTG 236

Qy         86 Asp----------------------------AlaGlyThrProGlyThrProGlyThr 95
              |||                            |||   :::|||      |||   |||
Db        237 GACCAGCTGCCGGAGCAGCCGCTCTTCCTCGCCGCCGCCTCGCCGCCCTGCCCATCTACT 296

Qy         96 ---ProValAlaThrProValTyrAlaThrProValGlyThrProMetGlyThrProSer 114
                 |||         |||         |||         |||         |||:::
Db        297 TCCCCGTCGCCGGAGCCCGCGGACGCGGCT---GCAGGAGCGAGTCGCTTCCAGCCCGCG 353

Qy        115 PheHisArgGlyThrProGlnTyrLysGlnArgSerGluLeuGlySerGlnGlyLysPro 134
                    |||   |||               |||      ||||||
Db        354 ------GCGGGACCGCCACCCCCGGGAGCGGCGAGTCGCTGCGGCTCCCACTCTGCCGAG 407

Qy        135 LeuHisArgArgArgArgSerGlnSerArgGluProGlyHisArgSerProSerArgGlu 154
              |||      |||   |||   :::|||   |||               |||
Db        408 CTGGCGGCCGCGCGGGACAGCGGCGCCCGGAGCCCCGCGGGGGCGGAGCCGCCCTCTGCA 467

Qy        155 ProSerAlaAspGlyArgProSerGluSerAlaGluProAspAspThrLeuGlyGlyGlu 174
                 :::      ||||||      |||:::   |||         ||||||   |||
Db        468 GCGGCCCCCTCCGGTCGAGAGATGGAGAATAAAGAA---------ACCCTCAAAGGA--- 515

Qy        175 TyrAlaTyrValTrpGlyThrAsnValAsnIleProAspValLeuArgAlaIleArgArg 194

Db        515 ------------------------------------------------------------ 515

Qy        195 PheLeuHisAsnTyrArgSerSerAlaHisAspLeuAsnSerLysTyrIleGlnIleIle 214
                 ||||||
Db        516 ---CTGCACAAGATGGAGGATCGCCCG------------------------------ 539

Qy        215 GluGluThrValGluArgGluGluAspThrLeuAsnIleAspMetSerAspIleTyrAsp 234
              ||||||   :::   ||||||:::         |||                     :::
Db        540 GAGGAGAGAATGATCCGGGAGAAG------CTCAAGGCGACCTGTATGCCGGCCTGGAAG 593

Qy        235 HisAspProAspLeuTyrAlaLysIleValArgTyrProLeuAspIleIleProLeu--- 253
              |||:::      |||         :::      |||   |||:::   :::   |||:::
Db        594 CACGAG---TGGTTGGAGAGGAGGAACAGGAGAGGCCCTGTGGTGGTGAAGCCAATCCCT 650

Qy        254 LeuAspThrGluCysGlnGluValAlaThrSerLeuLeuProThrPheGluLysHisIle 273
              :::      :::      ||||||                           ::::::
Db        651 ATTAAAGGAGATGGATCTGAAGTG-------------------------AATAACTTG 683

Qy        274 GluAlaArgProPheAsn--------------------LeuLysAlaSerValHisMet 286
                 |||   |||                                  |||
Db        684 GCAGCTGAGCCCCAGGGAGAGGGCCAGGCAGGTTCCGCTGCACCAGCCCCCAAGGGCCGA 743

Qy        287 ArgGluLeuAsnProSerAspIleAspLysLeuValSerValLysGlyMetValIleArg 306
              |||      :::|||                     |||||||||
Db        744 CGAAGCCCATCTCCTGGCAGCTCTCCGTCAGGGCGCTCGGTGAAG--------------- 788

Qy        307 CysSerSerIleIleProGluIleLysGlyAlaPhePheLysCysLeuValCysGlyHis 326
                                ||||||      |||         |||   :::
Db        789 ---------------CCGGAATCCCCAGGAGTAAGACGGAAACGAGTG----------- 821

Qy        327 SerProProLeuValThrValValLysGlyArgValGluGluProThrArgCysGlu--- 345
              ||||||      |||         ||||||:::         |||   |||
Db        822 TCCCCG------GTGCCTTTCCAGAGTGGCAGAATCACACCACCCCGAAGAGCCCCATCA 875

Qy        346 ---------------------LysProGluCysAlaAlaArgAsnAlaMetSerLeuIle 358
                                   ||||||      :::|||               ::::::
Db        876 CCGGATGGCTTCTCCCCGTACAGCCCAGAGGAGACGAGCCGCCGCGTGAACAAAGTGATG 935
```

```
Qy    359 HisAsnArg-CysThrPheAlaAsnLys---------GlnIleValArgLeuGlnGluTh 375
Db    936 AGAGCCAGGCTGTACCTGCTGCAGCAGATAGGACCCAACTCTTTCCTGATTGGAGGAGAC 995

Qy    375 rProAspAlaIleProGluGly-----------------------------GluThrPr 385
Db    996 AGTCCAGACAATAAATACCGGGTGTTTATTGGGCCACAGAACTGCAGCTGTGGGCGTGGA 1055

Qy    385 oHisThrVal---SerMetCysLeuTyrAsnThrMetValAspAlaValLysProGlyAs 404
Db   1056 GCATTCTGTATTCACCTCTTGTTTGTC-ATGCTCCGGGTGTTTCAGCTAGAACCCTCTGA 1114

Qy    404 pArgIleGluValThrGlyValPheLysAlaMetAlaValArgValGlyProAsnGlnAr 424
Db   1115 CCCCATGTTATGGAGAAAAACTTTAAAAAATTTCGAGGTT-------------------- 1154

Qy    424 gThrLeuArgAlaLeuTyrLysThrTyrIleAspCysValHisValLysLysSerAspAr 444
Db   1155 -------GAGAGTTTGTTCCAGAAATAC------------CACAGTAGGCGTAGCTCGAG 1195

Qy    444 gGlyArgLeuGlnThrGluAspProMetGlu---------------MetAspLysGluAs 459
Db   1196 AATCAAAGCTCCATCCCGGAACACCATCCAGAAGTTTGTGTCACGCATGTCAAATTCTCA 1255

Qy    459 nAspMetTyrAlaGlyTyrHisGluSerAspThrSerGluAlaAlaAsnGluAlaLysIl 479
Db   1256 CACACTG------------TCATCGTCTAGCACATCCACATCTAGTTCAGAAAACAGCAT 1303

Qy    479 eGlnLysLeuLysGluLeuSer---------LysLeuProGlyIleTyrAspArgLeuSe 496
Db   1304 CAAGGATGAAGAGGAGCAGATGTGTCCCATCTGCTTGCTGGGCATGCTGGATGAGGAGAG 1363

Qy    496 rArgSerLeuAlaPro-----------------------------------SerIleTr 504
Db   1364 CCTGACTGTGTGTGAAGATGGCTGCAGGAACAAGCTGCACCACCATTGCATGTCCATCTG 1423

Qy    504 p------GluLeuGluAspIleLysLysGlyLeuLeuCysGlnLeuPheGlyGlyLys-- 521
Db   1424 GGCGGAAGAGTGTAGAAGAAATAGAGAGCCTTTAATATGTCCCCTTTGTAGATCTAAGTG 1483

Qy    522 ------------------------------AlaLysLysIleProSerGlyAlaSerPh 531
Db   1484 GAGATCCCATGACTTCTACAGCCATGAGTTATCAAGCCCCGTGGAGTCCCCCGCCTCCCT 1543

Qy    531 eArg--------------------------------------------------------- 532
Db   1544 GCGAGCTGTCCAGCAGCCATCCTCCCCGCAGCAGCCCGTGGCCGGATCACAGCGGAGGAA 1603

Qy    533 -------GlyAspIleAsnValLeuLeuValGlyAsp---------ProGlyThrSerLy 547
Db   1604 TCAGGAGAGCAGTTTTAACCTTACTCATTTTGGAACCCAGCAGATTCCTTCCGCTTACAA 1663

Qy    547 sSer--------------------------------------------------------- 548
Db   1664 AGATTTGGCCGAGCCATGGATTCAGGTGTTTGGAATGGAACTCGTTGGCTGCTTATTCTC 1723

Qy    549 ------------------------GlnLeuLeuGlnTyrValHisLysIleAla----- 558
Db   1724 TAGAAACTGGAACGTAAGGGAAATGGCCCTTAGGCGTCTTTCCCACGACGTTAGTGGGGC 1783

Qy    559 ---------------------ProArgGlyIleTyrThrSerGlyArgGlySerSerAl 571
Db   1784 CCTGTTGTTGGCAAACGGGGAGAGCACTGGAAACTCTGGAGGCGGCAGTGGGGGCAGCTT 1843

Qy    571 aValGlyLeuThrAlaTyrValThrLysAspProGluThrArgGluThrValLeuGluSe 591
Db   1844 AAGCGCGGGAGCGGCCAGCGGGTCCTCCCAGCCCAGCATCTCAGGGGATGTGGTGGAGGC 1903

Qy    591 rGlyAlaLeuValLeuSerAspArgGlyIleCysCysIleAspGluPheAspLysMetSe 611
Db   1904 GTGCTGCAGTGTCCTGTCT---------ATAGTCTGCGCTGACCCTGTCTACAAAGTGTA 1954

Qy    611 rAspAsnAla------------ArgSerMetLeu---------------HisGluValMe 622
Db   1955 CGTTGCTGCTTTAAAAACATTGAGAGCCATGCTGGTATACACTCCTTGCCACAGTCTGGC 2014

Qy    622 tGluGln--------------------------------GlnThrValSerValAlaLy 631
Db   2015 AGAAAGAATCAAACTTCAGAGACTCCTCCGGCCAGTTGTAGACACTATCCTTGTCAAG-- 2072

Qy    631 sGlyGlyIleIleAlaSerLeuAsnAlaArgThrSerValLeuAla-------------- 646
Db   2073 ----------TGTGCAGATGCCAACAGCCGCACGAGTCAGCTGTCCATATCTACAGTGCT 2122

Qy    647 -------CysAlaAsnProSerGlySerArgTyrAsnAlaArg----------------- 658
Db   2123 GGAACTCTGCAAGGGCCAAGCAGGAGAGCTGGCGGTTGGGAGAGAAATACTTAAAGCTGG 2182

Qy    659 ------------------------------LeuSerValIle----------------- 662
Db   2183 GTCCATCGGGGTTGGTGGTGTCGATTACGTCTTAAGTTGTATCCTTGGAAACCAAGCTGA 2242

Qy    663 ----AspAsnIleGlnLeuProProThrLeuLeuSerArgPheAspLeuIle---TyrLe 680
Db   2243 ATCAAACAACTGGCAA---------GAACTGCTGGGTCGCCTCTGTCTTATAGACAGGTT 2293

Qy    680 uMetLeuAspLysProAspGluGlnAsnAspArgArgLeuAlaArgHisLeuValAlaLe 700
Db   2294 GCTGTTGGAATTTCCTGCTGAA---------------TTCTATCCTCATATTGTCAGTAC 2338

Qy    700 uHisTyrGluAsnTyrGluValSerLysGlnAspAlaLeuAspLeu------GlnThrLe 718
Db   2339 T---------------GATGTCTCACAAGCTGAGCCTGTTGAAATCAGGTACAAGAAGCT 2383

Qy    718 uThrAlaTyrIleThrTyrAlaArgGlnHisValHisProThrLeuSerAspGluAlaAl 738
Db   2384 GCTCTCCCTCTTAACCTTTGCCTTGCAATCCATTGACAATTCCCACTCG----------- 2432

Qy    738 aGluAspLeuIleAsnGlyTyrValGluMetArgGlnLysGlyAsnPheProGlySerSe 758
Db   2433 ----------ATGGTTGGCAAGCTCTCTCGGAGG------------ATATATCTGAGCTC 2470

Qy    758 rLysLysValIleThrAlaThrProArgGlnLeuGluSerMetIleArgIleSerGluAl 778
Db   2471 TGCCAGGATGGTGACCGCAGTGCCCGCTGTGTTTTCCAAGCTGGTAACCATGCTTAATGC 2530

Qy    778 aLeuAla---------------------ArgMetArgPheSerGluValValGluLysVa 791
Db   2531 TTCTGGCTCCACCCACTTCACCAGGATGCGCCGGCGTCTGATGGCTATCGCGGATGAGGT 2590

Qy    791 lAspAlaAlaGluAlaValArgLeu---------------LeuAsp-------------- 801
Db   2591 AGAAATTGCCGAGGTCATCCAGCTGGGTGTGGAGGACACTGTGGATGGGCATCAGGACAG 2650

Qy    802 ----------------------ValAlaLeuGlnGlnSerAlaThrAspHisAlaThrGl 814
Db   2651 CTTACAGGCCGTGGCCCCCACCAGCTGTCTAGAAAACAGCTCCCTTGAGCAC-------- 2702

Qy    814 yThrIleAspMetAspLeuIleThrThrGlyValSerAlaSerGluArgIleArgArgAl 834
Db   2703 -ACAGTCCATAGAGAGAAAACTGGAAAAGGACTAAGTGCTACGAGA-------------- 2747

Qy    834 aAsnLeuLeuAlaAlaLeuArgGluLeuIleAlaAspLysIleSerProGlySerSerSe 854
Db   2748 -------CTGAGTGCCAGCTCGGAGGACATTTCTGACAGACTGGCC---GGCGTCTCTGT 2797

Qy    854 rGlyLeuLysThrSerGlnLeuLeuGlu 863
Db   2798 AGGACTTCCCAGCTCAACAACAACAGAA 2825
```

RESULT 11
US-10-750-185-27656
; Sequence 27656, Application US/10750185
; Publication No. US20050260603A1
; GENERAL INFORMATION:

```
;  APPLICANT: MMI GENOMICS, INC.
;  APPLICANT:  DeNISE, Sue K.
;  APPLICANT:  KERR, Richard
;  APPLICANT:  ROSENFELD, David
;  APPLICANT:  HOLM, Tom
;  APPLICANT:  BATES, Stephen
;  APPLICANT:  FANTIN, Dennis
;  TITLE OF INVENTION: COMPOSITIONS FOR INFERRING BOVINE TRAITS
;  FILE REFERENCE: MMI1100-2
;  CURRENT APPLICATION NUMBER: US/10/750,185
;  CURRENT FILING DATE:  2003-12-31
;  PRIOR APPLICATION NUMBER: US 60/437,482
;  PRIOR FILING DATE: 2002-12-31
;  NUMBER OF SEQ ID NOS: 64922
;  SOFTWARE: PatentIN version 3.1
; SEQ ID NO 27656
;   LENGTH: 876
;   TYPE: DNA
;   ORGANISM: Bovine  19866881490050
US-10-750-185-27656

Alignment Scores:
Pred. No.:              0.425         Length:        876
Score:                  144.00        Matches:       47
Percent Similarity:     28.69%        Conservative:  23
Best Local Similarity:  19.26%        Mismatches:    23
Query Match:            3.12%         Indels:        151
DB:                     6             Gaps:          4

US-10-768-511-8 (1-901) x US-10-750-185-27656 (1-876)

Qy        324 CysGlyHisSerProProLeuValThrValValLysGlyArgValGluGluProThrArg 343
               |||   |||:::          |||    :::   :::||||||:::    ||||||
Db          3 TGTGCCCACACG---GCCCGCGTGGAAATAGACCGTGGCCGCATTGCCGAGCCCTGCGTG 59

Qy        344 CysGluLysProGluCysAlaAlaArgAsnAlaMetSerLeuIleHisAsnArgCysThr 363
               ||||||:::       |||          ::::::|||:::||||||||||||||
Db         60 TGCGAGCGC------TGCCATACCAGCCACAGCATGGCCCTCATCCACAACCGCTCTGTG 113

Qy        364 PheAlaAsnLysGlnIleValArg------------------------------------ 371
               |||::::::||||||:::||||||
Db        114 TTTTCCGACAAGCAGATGGTACGTGCCCCCTCCACGTGCCCTGTCCTTGTTCTGCCCTTG 173

Qy        371 ------------------------------------------------------------ 371

Db        174 CCGGGCCCTCTGAAGCCCAGTGTAACCTGCCCACTCATGTAACCCTAAAACCGCGCTTCC 233

Qy        371 ------------------------------------------------------------ 371

Db        234 TGACTGGGTTGACCTGGAAGGGAGTACCGTCCAGAATGAATTGGAAAAGCAGGCAGAGGG 293

Qy        371 ------------------------------------------------------------ 371

Db        294 TGATGCAGGCTCACCCACCGCCACCCTGATGTGGCCACCCCACATAGGGGAGCATTACTC 353

Qy        371 ------------------------------------------------------------ 371

Db        354 AACAATGCTCTCGTAATGGAGGTGAGACTGAGTTGCTGGAGACTAGTGTGTTGTGGTGAG 413

Qy        371 ------------------------------------------------------------ 371

Db        414 CATAGCCATCTTCCAGAACAACGTCCCTGGGGCAGTGGAGGTGCCAGGTATCTGGATGTC 473

Qy        372 ---------------------------------LeuGlnGluThrProAspAlaIlePr 380
                                               |||||||||:::|||:::    :::||
Db        474 CTGAGATGTCCATTCCCCATCCTTCCAGATCAAGCTCCAGGAGTCCCCTGAAGACATGCC 533

Qy        380 oGluGlyGluThrProHisThrValSerMetCysLeuTyrAsnThrMetValAspAlaVa 400
               |    |||:::||||||||||||||     :::         :::|||    :::||||||    ||
Db        534 CGCGGGGCAGACGCCGCACACGGTCGTCCTCTTTGCTCACAATGACCTCGTGGACAAGGT 593

Qy        400 lLysProGlyAspArgIleGluValThrGlyValPheLysAlaMetAlaValArgValGl 420
               |:::||||||||||||:::    |||||||||

Db        594 CCAGCCTGGGGACAGAGTGCACGTCACAGGT----------------------------- 624

Qy        420 yProAsnGlnArgThrLeuArgAlaLeuTyrLysThrTyrIleAspCysValHisValLy 440
                                                           :::    :::|||:::||
Db        625 -----------------------------------------GAGGGGCTCCACCTTAA 641

Qy        440 sLysSerAsp 443
               |    ||||||
Db        642 GGCCTCTGAC 651

RESULT 12
US-10-467-657-3451
; Sequence 3451, Application US/10467657
; Publication No. US20050260581A1
; GENERAL INFORMATION:
;  APPLICANT: CHIRON SpA
;  APPLICANT: FONTANA Maria Rita
;  APPLICANT: PIZZA Mariagrazia
;  APPLICANT: MASIGNANI Vega
;  APPLICANT: MONACI Elisabetta
;  TITLE OF INVENTION: GONOCOCCAL PROTEINS AND NUCLEIC ACIDS
;  FILE REFERENCE:
;  CURRENT APPLICATION NUMBER: US/10/467,657
;  CURRENT FILING DATE:  2003-08-11
;  PRIOR APPLICATION NUMBER: GB-0103424.8
;  PRIOR FILING DATE: 2001-02-12
;  NUMBER OF SEQ ID NOS: 9218
;  SOFTWARE: SeqWin99, version 1.04
; SEQ ID NO 3451
;   LENGTH: 2211
;   TYPE: DNA
;   ORGANISM: Neisseria gonorrhoeae
US-10-467-657-3451

Alignment Scores:
Pred. No.:              2.04          Length:        2211
Score:                  141.50        Matches:       157
Percent Similarity:     34.16%        Conservative:  104
Best Local Similarity:  20.55%        Mismatches:    292
Query Match:            3.07%         Indels:        215
DB:                     6             Gaps:          30

US-10-768-511-8 (1-901) x US-10-467-657-3451 (1-2211)

Qy         15 ProTyrProSer--------------GlnSerGluGlyValSerThrProLeuProGln 29
               |||   ||||||                     :::          ::::::|||:::
Db        233 CCGATGCCGTCGCCGCCACCCTGCTTGCCGACATCGGACGCTACGTCCCCGATTGGAACC 292

Qy         30 ValThrSerProSerPheAspAsnAlaAlaSerProValAlaGlyArgArgAlaValArg 49
                           ||||||      |||   ||||||   ||||||:::|||       |||
Db        293 TATTGGTTTCCGAGCGCTGCAACAGCACCGTCGCCGA--GCTGGTCAAAGGTGTGGACGA 350

Qy         50 GlnThrProThrSerAlaValArgArgArgGlyArgGluThrAspSerAlaArgArgArg 69
                  |||      :::|||        |||:::          :::   |||    |||
Db        351 AGTGCAGAAACTTACCCACTTCGCCCGGGTGGACAGCCTCGCCACGCCGGAAGAACGCGC 410

Qy         70 ArgSerArgSerArgSerLeuGlyAsnSerVal---------TyrSerSerProTyrAsp 86
                 :::                   |||:::               |||       |||
Db        411 ACAGCAAGCGGAAACCATGCGGAAAATGCTGCTGGCGATGGTTACCGACATCCGCGTCGT 470

Qy         87 AlaGlyThrProGlyThrProGlyThrProValAlaThrProValTyrAlaThrProVal 106
                       |||                                  ||||||   |||
Db        471 ATTAATCAAACTGGC-----------------------GATGCGTACGCGCACCCTGCA 506

Qy        107 GlyThrProMetGlyThrProSerPheHisArgGlyThrProGlnTyrLysGlnArgSer 126
                                                           :::|||||||||
Db        507 ---------------------------------------ATTTTTAAGCAACGCCCC 524

Qy        127 GluLeuGlySerGlnGlyLysProLeuHisArgArgArgArgSerGlnSerArgGluPro 146
                                 :::|||          ||||||||         |||   |||
Db        525 CGA--------------CAGCCCTGAAAAACGCGCCGTCGC------CAAAGAAACCCT 563

Qy        147 GlyHisArgSerProSerArgGluProSer--AlaAspGlyArgProSerGluSerAlaG 166
```

```
                  |||      |||:::|||:::|||    |||    |||        :::    |||::::
Db      564 CGACATCTTCGCCCCGCTCGCCAACCGCTTGGGCGTGTGGCAGCTCAAATGGCAGCTCGA 623

Qy      166 luProAspAspThrLeuGlyGlyGluTyrAlaTyrValTrpGlyThrAsnValAsnIleP 186
            ::                                  :::|||   :::   :::      |
Db      624 AG--------------------------------ATTTGGGCTTCCGCCATCAAGAA-C 649

Qy      186 roAspValLeuArgAlaIleArgArgPheLeuHisAsnTyrArgSerSerAlaHisAspL 206
            ||:::       |||   |||       |||          |||:::            :::
Db      650 CCGAAAAATACCGCGAAATCGCCCTGCTTTTGGACGAAAAACGCACCGAACGCCTCGAAT 709

Qy      206 euAsnSerLysTyrIleGlnIleIleGluGluThrValGluArgGluGluAspThrLeuA 226
                   ::::::   |||:::                |||   ::::::        ::::
Db      710 ACATCGAAAACTTCCTCGATATCCTGCGT------ACGGAACTCAAAAAATACAATATCC 763

Qy      226 snIleAspMetSerAspIleTyrAspHisAspProAspLeuTyrAlaLysIleValArgT 246
            ::  :::::::::           |||          :::|||   |||:::|||:::
Db      764 ACTTTGAAGTCGCCGGCCGTCCGAAACACATCTACTCCATTTACAAAAAAATGGTGAAGA 823

Qy      246 yrProLeuAspIleIleProLeuLeuAspThrGluCysGlnGluValAlaThrSerLeuL 266
                |||                |||   |||                   :::           :
Db      824 AAAAACTCAGCTTCGACGGCCTGTTCGACATCCGCGCCGTGCGGATTCTGGTCGATACCG 883

Qy      266 euProThrPheGluLysHisIleGluAlaArgProPheAsnLeuLysAlaSerValHisM 286
            ::|||           |||         ||||||      |||   |||       |||
Db      884 TCCCCGAGTGTTA-CACCACGCTGGGCATCGTCCACAGCCTCTGGCAGCC------CATT 936

Qy      286 etArg-GluLeuAsnProSerAspIleAspLysLeuValSerValLysGly-MetValIl 305
              ||| |||    :::              |||   :::   :::   ||||||    |||||
Db      937 CCCGGCGAGTTCGAC------------GACTACATCGCCAACCCCAAAGGCAACGGTTAT 984

Qy      305 eArg---CysSerSerIleIleProGluIleLysGlyAlaPhePheLysCys-------- 321
            |:::       |||:::                              :::||||||
Db      985 AAAAGTTTGCACACCGTCATCGTCGGCCCGGAAGACAAAGGTGTGGAAGTGCAAATCCGC 1044

Qy      322 ----LeuValCysGlyHis---------------SerProProLeu---ValThrValVa 334
                      :::|||    :::                  |||||||||      ||||||
Db     1045 ACCTTCGATATGCACCAATTCAACGAATTCGGTGTCGCCGCCCACTGGCGTTACAAAGAA 1104

Qy      334 lLysGlyArgValGluGluProThrArgCysGluLysProGluCysAla----------- 350
                  :::           ||||||    :::   |||   ||||||
Db     1105 GGCGGCAAAGGCGATTCCGCCTACGAACAAAAAATCGCCTGGTTGCGCCAACTCTTGGAC 1164

Qy      351 ----------------------AlaArgAsnAlaMetSerLeuIleHisAsnArgCysTh 363
                                  |||                |||            ::::::   ::
Db     1165 TGGCGCGAAAATATGGCGGAAAGCGGCAAGGAAGACCTCGCCGCCGCCTTCAAAACCGAG 1224

Qy      363 rPheAlaAsnLysGlnIleValArgLeuGlnGluThrProAsp------------AlaIl 379
            :|||    |||    |||            |||      ||||||                 ::::
Db     1225 CTTTTC-AACGACACGATTTATGTTTTG------ACCCCGCACGGCAAAGTCCTCTCTCT 1277

Qy      379 eProGluGlyGluThrProHisThrValSerMetCysLeuTyrAsnThrMetValAsp-- 398
            :|||   |||      ||||||             :::   |||:::::::::::::     |||
Db     1278 GCCAACGGGCGCAACCCCCATCGACTTCGCCTACGCCCTGCACAGCAGCATCGGCGACCG 1337

Qy      399 ---------------------------------------------AlaValLysProGl 403
                                                           ::::::    ||
Db     1338 CTGCCGGGGCGCGAAAGTCGAAGGGCAGATTGTGCCGCTGTCCACCCCGCTCGAAAACGG 1397

Qy      403 yAspArgIleGluValThrGlyValPheLysAlaMetAlaValArgValGlyProAsnGl 423
             |   |||:::|||:::                                 :::
Db     1398 ACAGCGCGTCGAAATC------------------------ATTACCGCCAAAGAAGGGCA 1433

Qy      423 nArgThrLeuArgAlaLeuTyrLysThrTyrIleAspCys-------------------- 436
              ::::::          ||||||:::     ::::::
Db     1434 TCCTTCCGTCAACTGGCTTTACGAAGGCTGGGTCAAATCCGGCAAGGCCATCGGCAAAAT 1493

Qy      437 ----ValHisValLysLysSerAspArgGlyArgLeuGlnThrGluAspProMetGluMe 455
                :::::::::::       :::       ::::::   |||         ::::::::
Db     1494 CCGCGCCTACATCCGCCAGCAAAACGCCGACACCGTGCGCGAAGAAGGCCGTGTCCAACT 1553

Qy      455 tAspLysGlu-------------------------------------------------- 458
             :||||||:::
Db     1554 CGACAAGCAGCTTGCCAAACTCACGCCCAAACCCAACCTGCAAGAGCTTGCCGAAAATCT 1613

Qy      459 ----------------AsnAspMetTyrAlaGlyTyrHisGluSerAspThrSerGluAl 473
                               |||:::|||                   :::   :::   |||
Db     1614 CGGCTACAAAAAGCCAGAAGACCTCTACACCGCCGTCGGACAAGGCGAAATTTCCAACCG 1673

Qy      473 aAlaAsnGluAlaLysIleGlnLysLeuLysGluLeuSerLysLeuProGlyIleTyrAs 493
              |||   :::                 |||   |||            :::|||
Db     1674 CGCCATCCAAAAAGCCTGCGGCACGCTGAACGAACCGCCGCCCGTGCCC----------- 1722

Qy      493 pArgLeuSerArgSerLeuAlaProSerIleTrpGluLeuGluAspIleLysLysGlyLe 513
                          |||   :::|||   :::                |||||||||||
Db     1723 -------------GTCAGCGCAACCACCATCGTCAAACAGTCCAAAATCAAAAAAGGT-- 1767

Qy      513 uLeuCysGlnLeuPheGlyGlyLysAlaLysLysIleProSerGlyAlaSerPheArgGl 533
                                                                          ||
Db     1768 ----------------------------------------------------------GG 1769

Qy      533 yAspIleAsnValLeuLeuValGlyAspProGlyThrSerLysSerGlnLeuLeuGlnTy 553
            |           ||||||:::   |||:::   |||               :::   |||
Db     1770 CAAAACCGGCGTGCTCATCGACGGCGAAGACGGCTTGATGACCACG------CTTGCCAA 1823

Qy      553 rValHisLysIleAlaPro------------ArgGlyIleTyrThrSerGlyArgGlySe 569
               |||   ||||||                 |||      |||      ||||||
Db     1824 ATGCTGCAAACCCGCGCCGCCCGACGATATTGCCGGCTTCGTTACCCGCGAGCGCGGCAT 1883

Qy      569 rSerAlaValGlyLeuThrAlaTyrValThrLysAspProGluThrArg----------- 585
              |||         |||                  |||      |||
Db     1884 TTCCGTCCACCGCAAAACCTGC---------------CCCTCTTTCCGACACCTTGCCGA 1928

Qy      586 ----------GluThrValLeuGluSerGlyAlaLeuValLeuSerAspArgGlyIleCy 602
                          |||   ||||||::::::             |||   :::        :::
Db     1929 ACACGCGCCCGAAAAAGTACTGGACGCAAGTTGGGCGGCGTTGCAGGAAGGGCAAGTGTT 1988

Qy      602 sCysIleAspGluPheAspLysMetSerAspAsnAlaArgSerMetLeuHisGluValMe 622
                :::|||         :::         |||   :::      :::|||   :::|||
Db     1989 CGCCGTCGATATCGAAATCCGCGCCCAAGACCGCTCC---GGGCTTTTGCGCGACGTATC 2045

Qy      622 tGluGlnGlnThrValSerValAlaLysGlyGlyIleIleAlaSerLeu-AsnAlaArgT 642
             :::                                        :::|||    |||
Db     2046 CGAC---------------------------------------GCGCTCGCCCGCCACAA 2066

Qy      642 hrSerValLeuAlaCysAlaAsnProSer-----------------------GlySerA 654
              |||   |||         ||||||                          |||||||
Db     2067 ACTCAACGTTACCGCCGTGCAAACCCAGTCCCGCGACTTGGAAGCCAGCATGAGGTTCAC 2126

Qy      654 rgTyrAsnAlaArgLeuSerValIleAspAsnIleGlnLeuProProThrLeuLeuSerA 674
            ||    :::      |||                  |||   :::   :::|||:
Db     2127 GCTCGAAGTCAAACAAGTCAACGACCTCCCGCGCGTCCTCGCCGGCCTCGGCGATGTCAA 2186

Qy      674 rg 674
            ::
Db     2187 AG 2188
```

RESULT 13
US-11-077-716-1/c
; Sequence 1, Application US/11077716
; Publication No. US20050260180A1
; GENERAL INFORMATION:
; APPLICANT: WEI, LISA
; APPLICANT: RASMUSSEN, HENRIK S
; TITLE OF INVENTION: MATERIALS AND METHODS FOR TREATING VASCULAR LEAKAGE IN THE EYE
; FILE REFERENCE: 233988
; CURRENT APPLICATION NUMBER: US/11/077,716
; CURRENT FILING DATE: 2005-03-11
; PRIOR APPLICATION NUMBER: US 60/552,768
; PRIOR FILING DATE: 2004-03-12
; NUMBER OF SEQ ID NOS: 3
; SOFTWARE: PatentIn version 3.2
; SEQ ID NO 1
; LENGTH: 33014
; TYPE: DNA
; ORGANISM: Artificial

```
;   FEATURE:
;   OTHER INFORMATION: Synthetic construct
US-11-077-716-1

Alignment Scores:
Pred. No.:              127          Length:         33014
Score:                  137.00       Matches:        131
Percent Similarity:     35.59%       Conservative:   74
Best Local Similarity:  22.74%       Mismatches:     250
Query Match:            2.97%        Indels:         125
DB:                     7            Gaps:           20

US-10-768-511-8 (1-901) x US-11-077-716-1 (1-33014)

Qy        37 AsnAlaAlaSerProValAlaGly------------ArgArgAlaValArgGlnThrPro 52
Db      7995 CACGCAGAGGCGCCAGATTCAGGAGATCAACCGCCGCGTCGTCGCGTTCGCCAGCAACCT 7936

Qy        53 ThrSerAlaValArgArgArgGlyArgGluThrAspSerAlaArgArgArgArgSerArg 72
Db      7935 ACGCGCGCAGCACCAGCTCCTGCCCGC------------GCGCGGCGCCGACGTGCCCCT 7888

Qy        73 SerArgSerLeuGlyAsnSerValTyrSerSerProTyrAspAlaGlyThrProGlyThr 92
Db      7887 GCCCCCTCTCCCGGCGGGTCC-----------------------GGAGCCCCCCCTACC 7852

Qy        93 ProGlyThrProValAlaThrProValTyrAlaThrProValGlyThrProMetGlyThr 112
Db      7851 TCCGGGGGCTCGCCCGCGTCACCGCTTTTAGAT---------GCATCATCCAAGGACACC 7801

Qy       113 ProSerPheHisArgGlyThrProGlnTyrLysGlnArgSerGluLeuGlySerGlnGly 132
Db      7800 CCCGCGGCCCACCGCCC-GCCGCGCGGTACCGTAGTCGCGCC---------GCGGGGATG 7751

Qy       133 LysProLeuHisArgArgArgArgSerGlnSerArgGluProGlyHis------------ 148
Db      7750 CGGCCTCTTGCAAGCCATCGACGCCGCCACCAACCAGCCCCTGGAAATTAGGTATCACCT 7691

Qy       149 ---ArgSerProSerArgGluProSerAlaAspGlyArgProSerGluSerAlaGluPro 167
Db      7690 GGATCTAGCCCGCGCCCTGACCCGTCTATGCGAGGTAAACCTGCAGGAGCTCCCGCCTGA 7631

Qy       168 AspAspThrLeuGlyGlyGluTyrAlaTyrValTrpGlyThrAsnValAsnIleProAsp 187
Db      7630 CCTGACGCCGCGGGAGCT-----------------------------------CCAGAC 7607

Qy       188 -Val------------LeuArgAlaIleArgArgPheLeuHisAsnTyrArgSerSerAl 203
Db      7606 CATGGACAGCTCCCATCTGCGCGATGTTGTCATCAAGCTCCGACCGCCGCGCGCGGACAT 7547

Qy       203 aHisAspLeuAsnSerLysTyrIleGlnIleIleGluGluThrValGluArgGluGluAs 223
Db      7546 CTGGACTTTGGGCTCGCGCGGCGTGGTGGTCCGATCCACCGTAACTCCCCTCGAGCAGCC 7487

Qy       223 pThrLeuAsnIleAspMetSerAspIleTyrAspHisAspProAspLeuTyrAlaLysIl 243
Db      7486 AGACGGTCAAGGACAAGCAGCCGAAGTAGAAGACCACCAGCCAAACCCGCCAGGCGAGGG 7427

Qy       243 eValArgTyrProLeu---------------------------------------AspIl 250
Db      7426 GCTCAAATTCCCACTCTGCTTCCTTGTGCGCGGTCGTCAGGTCAACCTCGTGCAGGATGT 7367

Qy       250 eIleProLeuLeuAspThrGluCysGlnGluValAlaThrSerLeuLeuProThrPheGl 270
Db      7366 ACAGCCCGTG------CACCGCTGCCAGTACTGCGCACGTTTTCACAAAAGCCAGCACGA 7313

Qy       270 uLysHisIleGluAlaArgProPhe---------------AsnLeuLysAlaSerValHi 285
Db      7312 GTGTTCGGCCCGTCGCAGGGACTTCTACTTTCACCACATCAATAGCCACTCCTCCAATTG 7253

Qy       285 sMetArgGluLeuAsnProSerAspIleAspLysLeuValSerValLysGlyMetValIl 305
Db      7252 GTGGCGGGAGATCCAGTTCTTCCCGATCGGCTCGCATCCTCGCACCGAGCGTCTCTTTGT 7193

Qy       305 eArgCysSerSerIleIleProGluIleLysGlyAlaPhePheLysCysLeuVal----- 323
Db      7192 CACCTACGATGTAGAGACCTATACTTGGATGGGGGCCTTTGGGAAGCAGCTCGTGCCCTT 7133

Qy       324 ----------------CysGlyHisSerProProLeuValThrValValLysGlyArgVa 338
Db      7132 CATGCTGGTCATGAAGTTCGGCGGAGATGAGCCTCTAGTGACTGCCGCGCGAGACCT-AG 7074

Qy       338 lGluGluProThrArgCysGluLysProGluCysAlaAlaArgAsnAlaMetSerLeuIl 358
Db      7073 CCGCGAACCTTGG---------ATGGGACCGCTGGGAACAAGACCCGCTTACCTTCTACT 7023

Qy       358 eHisAsnArgCysThrPheAlaAsnLysGlnIleValArgLeuGlnGluThrProAspAl 378
Db      7022 GCATCACCCCAGAAAAAATGGCCATAGGTCGCCAGTTTAGGACCTTTCGCGACCACCTGC 6963

Qy       378 aIleProGluGlyGluThrProHisThrValSerMetCysLeuTyrAsnThrMetValAs 398
Db      6962 AAATGCTAATGGCCCGTGACCT-------------------------------------- 6941

Qy       398 pAlaValLysProGlyAspArgIleGluValThrGlyValPheLysAlaMetAlaValAr 418
Db      6940 ----------------------GTGGAGCTCATTCGTCGCTTCCAACCCTCATCTTGCAG 6903

Qy       418 gValGlyProAsnGlnArgThrLeuArgAlaLeuTyrLysThrTyrIleAspCysValHi 438
Db      6902 ACTGGGCCCTTT-CAGAGCACGGGCTCAGCT--------------------------CC 6871

Qy       438 sValLysLysSerAspArgGlyArgLeuGlnThrGluAspProMetGluMetAspLysGl 458
Db      6870 CCTGAAGAGCTCACCTACGAGGAACTTAAAAAATTGCCTTCCATCAAGGGCATCCCGCGC 6811

Qy       458 uAsnAspMetTyrAlaGlyTyrHisGluSerAspThrSerGluAlaAlaAsnGluAlaLy 478
Db      6810 TTCTTGGAACTTTACATTGTGGGCCACAACATCAACGGCTTTGACGAGATCGTG-CTCGC 6752

Qy       478 sIleGlnLysLeuLysGluLeuSerLysLeuProGlyIleTyrAspArgLeuSerArgSe 498
Db      6751 CGCCCAGGTAATTAACAACCGTTCCGAGGTGCCGGGACCCTTC---CGCATCACACGCAA 6695

Qy       498 rLeuAlaProSerIleTrpGluLeu-GluAspIleLysLysGlyLeuLeuCysGlnLeuP 518
Db      6694 CTTTATGCCTCGCGCGGGAAAGATACTCTTCAACGATGTCACCTTCGCCCTGCCAAATC- 6636

Qy       518 heGlyGlyLysAlaLysLysIleProSerGlyAla---------------SerPheArgG 533
Db      6635 -----------------CGCGTTCCAAAAAGCGCACGGACTTTTGCTCTGGGAGCAGGG 6593

Qy       533 lyAspIleAsnValLeuLeuValGlyAspProGlyThrSerLysSerGlnLeuLeuGlnT 553
Db      6592 CGGATGCGACGACA---CTGACTTCAAATACCAGTACCTCAAAGTCATGGTCAGGGACAC 6536

Qy       553 yrValHisLysIleAlaProArgGlyIleTyrThrSerGlyArg 567
Db      6535 CTTTGCGCTCACCCACACCTCGC------------TCCGGAAGG 6504

RESULT 14
US-10-623-155-479/c
; Sequence 479, Application US/10623155
; Publication No. US20050261166A1
; GENERAL INFORMATION:
;  APPLICANT: Wang, Tongtong
;  APPLICANT:  Peckham, David W.
;  APPLICANT:  Retter, Marc W.
;  APPLICANT:  Fanger, Gary R.
;  TITLE OF INVENTION: COMPOSITIONS AND METHODS FOR THE THERAPY
;  TITLE OF INVENTION:  AND DIAGNOSIS OF LUNG CANCER
;  FILE REFERENCE: 210121.455C20
;  CURRENT APPLICATION NUMBER: US/10/623,155
;  CURRENT FILING DATE:  2003-07-17
;  NUMBER OF SEQ ID NOS: 560
;  SOFTWARE: FastSEQ for Windows Version 4.0
```

```
; SEQ ID NO 479
;   LENGTH: 34555
;   TYPE: DNA
;   ORGANISM: Homo sapiens
;   FEATURE:
;   NAME/KEY: misc_feature
;   LOCATION: 34543,34544,34545,34546,34547
;   OTHER INFORMATION: n = A,T,C or G
US-10-623-155-479

Alignment Scores:
Pred. No.:              135          Length:         34555
Score:                  137.00       Matches:        131
Percent Similarity:     35.59%       Conservative:   74
Best Local Similarity:  22.74%       Mismatches:     250
Query Match:            2.97%        Indels:         125
DB:                     6            Gaps:           20

US-10-768-511-8 (1-901) x US-10-623-155-479 (1-34555)

Qy        37 AsnAlaAlaSerProValAlaGly------------ArgArgAlaValArgGlnThrPro 52
Db     10022 CACGCAGAGGCGCCAGATTCAGGAGATCAACCGCCGCGTCGTCGCGTTCGCCAGCAACCT 9963

Qy        53 ThrSerAlaValArgArgArgGlyArgGluThrAspSerAlaArgArgArgArgSerArg 72
Db      9962 ACGCGCGCAGCACCAGCTCCTGCCCGC------------GCGCGGCGCCGACGTGCCCCT 9915

Qy        73 SerArgSerLeuGlyAsnSerValTyrSerSerProTyrAspAlaGlyThrProGlyThr 92
Db      9914 GCCCCCTCTCCCGGCGGGTCC------------------------GGAGCCCCCCCTACC 9879

Qy        93 ProGlyThrProValAlaThrProValTyrAlaThrProValGlyThrProMetGlyThr 112
Db      9878 TCCGGGGGCTCGCCCGCGTCACCGCTTTTAGAT---------GCATCATCCAAGGACACC 9828

Qy       113 ProSerPheHisArgGlyThrProGlnTyrLysGlnArgSerGluLeuGlySerGlnGly 132
Db      9827 CCCGCGGCCCACCGCCC-GCCGCGCGGTACCGTAGTCGCGCC---------GCGGGGATG 9778

Qy       133 LysProLeuHisArgArgArgArgSerGlnSerArgGluProGlyHis------------ 148
Db      9777 CGGCCTCTTGCAAGCCATCGACGCCGCCACCAACCAGCCCCTGGAAATTAGGTATCACCT 9718

Qy       149 ---ArgSerProSerArgGluProSerAlaAspGlyArgProSerGluSerAlaGluPro 167
Db      9717 GGATCTAGCCCGCGCCCTGACCCGTCTATGCGAGGTAAACCTGCAGGAGCTCCCGCCTGA 9658

Qy       168 AspAspThrLeuGlyGlyGluTyrAlaTyrValTrpGlyThrAsnValAsnIleProAsp 187
Db      9657 CCTGACGCCGCGGGAGCT------------------------------------CCAGAC 9634

Qy       188 -Val------------LeuArgAlaIleArgArgPheLeuHisAsnTyrArgSerSerAl 203
Db      9633 CATGGACAGCTCCCATCTGCGCGATGTTGTCATCAAGCTCCGACCGCCGCGCGCGGACAT 9574

Qy       203 aHisAspLeuAsnSerLysTyrIleGlnIleIleGluGluThrValGluArgGluGluAs 223
Db      9573 CTGGACTTTGGGCTCGCGCGGCGTGGTGGTCCGATCCACCGTAACTCCCCTCGAGCAGCC 9514

Qy       223 pThrLeuAsnIleAspMetSerAspIleTyrAspHisAspProAspLeuTyrAlaLysIl 243
Db      9513 AGACGGTCAAGGACAAGCAGCCGAAGTAGAAGACCACCAGCCAAACCCGCCAGGCGAGGG 9454

Qy       243 eValArgTyrProLeu---------------------------------------AspIl 250
Db      9453 GCTCAAATTCCCACTCTGCTTCCTTGTGCGCGGTCGTCAGGTCAACCTCGTGCAGGATGT 9394

Qy       250 eIleProLeuLeuAspThrGluCysGlnGluValAlaThrSerLeuLeuProThrPheGl 270
Db      9393 ACAGCCCGTG------CACCGCTGCCAGTACTGCGCACGTTTTTACAAAAGCCAGCACGA 9340

Qy       270 uLysHisIleGluAlaArgProPhe---------------AsnLeuLysAlaSerValHi 285
Db      9339 GTGTTCGGCCCGTCGCAGGGACTTCTACTTTCACCACATCAATAGCCACTCCTCCAATTG 9280

Qy       285 sMetArgGluLeuAsnProSerAspIleAspLysLeuValSerValLysGlyMetValIl 305
Db      9279 GTGGCGGGAGATCCAGTTCTTCCCGATCGGCTCGCATCCTCGCACCGAGCGTCTCTTGT 9220

Qy       305 eArgCysSerSerIleIleProGluIleLysGlyAlaPhePheLysCysLeuVal----- 323
Db      9219 CACCTACGATGTAGAGACCTATACTTGGATGGGGGCCTTTGGGAAGCAGCTCGTGCCCTT 9160

Qy       324 ----------------CysGlyHisSerProProLeuValThrValValLysGlyArgVa 338
Db      9159 CATGCTGGTCATGAAGTTCGGCGGAGATGAGCCTCTAGTGACTGCCGCGCGAGACCT-AG 9101

Qy       338 lGluGluProThrArgCysGluLysProGluCysAlaAlaArgAsnAlaMetSerLeuIl 358
Db      9100 CCGCGAACCTTGG---------ATGGGACCGCTGGGAACAAGACCCGCTTACCTTCTACT 9050

Qy       358 eHisAsnArgCysThrPheAlaAsnLysGlnIleValArgLeuGlnGluThrProAspAl 378
Db      9049 GCATCACCCCAGAAAAAATGGCCATAGGTCGCCAGTTTAGGACCTTTCGCGACCACCTGC 8990

Qy       378 aIleProGluGlyGluThrProHisThrValSerMetCysLeuTyrAsnThrMetValAs 398
Db      8989 AAATGCTAATGGCCCGTGACCT-------------------------------------- 8968

Qy       398 pAlaValLysProGlyAspArgIleGluValThrGlyValPheLysAlaMetAlaValAr 418
Db      8967 ----------------------GTGGAGCTCATTCGTCGCTTCCAACCCTCATCTTGCAG 8930

Qy       418 gValGlyProAsnGlnArgThrLeuArgAlaLeuTyrLysThrTyrIleAspCysValHi 438
Db      8929 ACTGGGCCCTTT-CAGAGCACGGGCTCAGCT-------------------------CC 8898

Qy       438 sValLysLysSerAspArgGlyArgLeuGlnThrGluAspProMetGluMetAspLysGl 458
Db      8897 CCTGAAGAGCTCACCTACGAGGAACTTAAAAAATTGCCTTCCATCAAGGGCATCCCGCGC 8838

Qy       458 uAsnAspMetTyrAlaGlyTyrHisGluSerAspThrSerGluAlaAlaAsnGluAlaLy 478
Db      8837 TTCTTGGAACTTTACATTGTGGGCCACAACATCAACGGCTTTGACGAGATCGTG-CTCGC 8779

Qy       478 sIleGlnLysLeuLysGluLeuSerLysLeuProGlyIleTyrAspArgLeuSerArgSe 498
Db      8778 CGCCCAGGTAATTAACAACCGTTCCGAGGTGCCGGGACCCTTC---CGCATCACACGCAA 8722

Qy       498 rLeuAlaProSerIleTrpGluLeu-GluAspIleLysLysGlyLeuLeuCysGlnLeuP 518
Db      8721 CTTTATGCCTCGCGCGGGAAAGATACTCTTCAACGATGTCACCTTCGCCCTGCCAAATC- 8663

Qy       518 heGlyGlyLysAlaLysLysIleProSerGlyAla---------------SerPheArgG 533
Db      8662 -----------------CGCGTTCCAAAAAGCGCACGGACTTTTGCTCTGGGAGCAGGG 8620

Qy       533 lyAspIleAsnValLeuLeuValGlyAspProGlyThrSerLysSerGlnLeuLeuGlnT 553
Db      8619 CGGATGCGACGACA---CTGACTTCAAATACCAGTACCTCAAAGTCATGGTCAGGGACAC 8563

Qy       553 yrValHisLysIleAlaProArgGlyIleTyrThrSerGlyArg 567
Db      8562 CTTTGCGCTCACCCACACCTCGC------------TCCGGAAGG 8531

RESULT 15
US-11-074-176-47
; Sequence 47, Application US/11074176
; Publication No. US20050250135A1
; GENERAL INFORMATION:
;  APPLICANT: Klaenhammer, Todd R.
;  APPLICANT: Russell, William M.
;  APPLICANT: Altermann, Eric
;  APPLICANT: McAuliffe, Olivia
;  APPLICANT: Peril, Andrea Azcarate
```

```
;  TITLE OF INVENTION: Nucleic Acid Sequences Encoding
;  TITLE OF INVENTION:  Stress-Related Proteins and Uses Therefore
;  FILE REFERENCE: 5051-694
;  CURRENT APPLICATION NUMBER: US/11/074,176
;  CURRENT FILING DATE:  2005-03-07
;  PRIOR APPLICATION NUMBER: 60/551,161
;  PRIOR FILING DATE: 2004-03-08
;  NUMBER OF SEQ ID NOS: 381
;  SOFTWARE: FastSEQ for Windows Version 4.0
; SEQ ID NO 47
;   LENGTH: 1260
;   TYPE: DNA
;   ORGANISM: Lactobacillus acidophilus
;   FEATURE:
;   NAME/KEY: CDS
;   LOCATION: (1)...(1260)
;   FEATURE:
;   NAME/KEY: misc_feature
;   LOCATION: (0)...(0)
;   OTHER INFORMATION: ORF 847; ClpX - ATP-dependent protease/ATP-binding
;   OTHER INFORMATION: subunit
US-11-074-176-47

Alignment Scores:
Pred. No.:              7.74          Length:        1260
Score:                  129.00        Matches:       103
Percent Similarity:     38.79%        Conservative:  77
Best Local Similarity:  22.20%        Mismatches:    164
Query Match:            2.80%         Indels:        120
DB:                     7             Gaps:          23

US-10-768-511-8 (1-901) x US-11-074-176-47 (1-1260)

Qy     433 TyrIleAspCysValHisValLysLysSerAspArgGlyArgLeuGlnThrGluAspPro 452
             :::   |||           :::|||   :::     |||:::   |||
Db      13 TTTACAGATCAAGAAGAGATTAAATGTGCTTTTTGTGGTAAAACTCAA------------ 60

Qy     453 MetGluMetAspLysGluAsnAspMetTyrAlaGly---------TyrHisGluSerAsp 469
                      |||:::         |||   |||||         |||   ::::::
Db      61 ---------GACCAAGTTAAAAAAATGATTGCCGGAAACGGTGTATATATTTGTAATGAA 111

Qy     470 ThrSerGluAlaAla---------------AsnGluAlaLysIleGlnLysLeuLysGlu 484
              :::   |||               :::|||   :::         |||||
Db     112 TGTGTTGACTTGGCTAAAAAAATTATTGACGATGAATTACGTGCAGATTCACTAAAAACT 171

Qy     485 LeuSerLysLeuProGlyIleTyrAspArgLeuSerArgSerLeuAlaProSerIleTrp 504
                 |||:::||||||
Db     172 GCAAGTGAACTGCCT--------------------------------------------- 186

Qy     505 GluLeuGluAspIleLysLysGlyLeuLeuCysGlnLeuPheGly---GlyLysAlaLys 523
             :::      :::|||||||||   |||         :::   |||      :::||||||
Db     187 AAACCAGTTGAAATTAAAAAACAACTTGATCAATATGTAATTGGTCAAGATCGTGCTAAA 246

Qy     524 LysIleProSerGlyAlaSerPhe--------------------ArgGlyAspIle--- 535
             |||:::   |||   |||   :::                   :::   |||:::
Db     247 AAAGTACTATCAGTTGCTGTTTATAATCACTACAAGAGAATCAGCCAAATGGATGTAGAT 306

Qy     536 -----------------------AsnValLeuLeuValGlyAspProGlyThrSerLys 547
                                  |||:::   ::::::|||      |||:::   |||
Db     307 TCATCAACAGAATTACAAAAGTCTAATATTGCAATGATTGGTCCTACTGGATCAGGTAAG 366

Qy     548 SerGlnLeuLeuGlnTyrValHisLysIle------AlaProArgGlyIleTyrThrSer 565
             :::   |||   |||   :::   :::|||         |||      |||
Db     367 ACTTATTTAGCTCAAACACTTGCACGTATTTTAAACGTTCCATTTGCAATT--------- 417

Qy     566 GlyArgGlySerSerAlaValGlyLeuThr------AlaTyrValThrLysAspProGlu 583
                      :::   |||      |||||         |||||   :::|||   |||
Db     418 ---------GCTGATGCAACCACGTTAACTGAAGCTGGATATGTTGGTGAAGATGTAGAA 468

Qy     584 ThrArgGluThrValLeuGluSerGlyAla---------LeuValLeuSerAspArgGly 600
                        |||         |||         |||      :::   ||||||
Db     469 AACATTTTGCTTAAATTATTACAAAATGCTGATTACGATCTTGAACGAGCACAACGTGGA 528

Qy     601 IleCysCysIleAspGluPheAspLysMetSerAspAsnAla----------------- 614
             |||         |||||||||   ||||||:::|||         :::
Db     529 ATTATTTATATTGATGAAATTGATAAGATTTCTAAAAAATCTGAAAACGTATCAATTACA 588

Qy     615 ArgSerMetLeuHisGluValMetGluGlnGlnThrValSerValAlaLysGlyGlyIle 634
             |||   :::         |||   ::::::|||         :::   :::   :::|||
Db     589 CGTGATGTTTCAGGTGAAGGTGTGCAACAATCGCTTTTGAAGATTTTGGAAGGT------ 642

Qy     635 IleAlaSerLeuAsnAlaArgThrSerValLeuAlaCysAlaAsnProSerGlySerArg 654
                                         :::|||         |||   |||   |||
Db     643 -------------------------ACTATTGCTTCTGTGCCTCCTCAAGGTGGACGA 675

Qy     655 TyrAsnAlaArgLeuSerValIleAspAsnIleGlnLeuProProThrLeuLeuSerArg 674
                :::   :::      :::|||                           :::
Db     676 AAGCATCCGCAACAAGAAATGATAAAG-----------------------ATGGATACT 711

Qy     675 PheAspLeuIleTyrLeuMet---------LeuAspLysProAspGluGlnAsnAspArg 691
                :::::::::::::::::                |||         ::::::
Db     712 ACTAATATTCTCTTTATTGTAGGTGGAGCATTTGATGGAATTGAACAGATTGTTAAAAGC 771

Qy     692 ArgLeuAlaArgHisLeuValAlaLeuHisTyrGluAsnTyrGluValSerLysGlnAsp 711
             ||||||   :::      :::                ||||||   ||||||:::|||   |||
Db     772 CGTTTAGGTAAGAAAACTATTGGTTTTGGTGCTGAAAAC---GAAGTTAACAAGGTAGAT 828

Qy     712 AlaLeuAspLeuGlnThrLeuThrAlaTyrIleThrTyrAla----------------- 725
             |||   |||               |||   ::::::|||   |||
Db     829 GCTGATGAT---------TGGACTCGTCATTTAACTACTGCTGATTTAGTTAAATTCGGT 879

Qy     726 --------------------ArgGlnHisValHisProThrLeuSerAspGluAlaAla 738
                               |||      :::      ||||||
Db     880 ATGATTCCAGAATTCATTGGTAGAATTCCAATTATTACCACGCTTGATAAACTTGATAAT 939

Qy     739 GluAspLeuIleAsnGlyTyrValGluMetArgGlnLysGlyAsnPheProGlySerSer 758
             :::||||||:::                  |||         |||
Db     940 AAAGATTTAGTTCGCGTATTAACTGAG------CCAAAGAATGCTCTTGTTAAGCAATAT 993

Qy     759 LysLysValIleThrAlaThrProArgGlnLeuGluSerMetIleArgIleSerGluAla 778
             ||||||:::::::::                  :::|||:::                  :::|||
Db     994 AAGAAACTTCTATCATTAGATGGTGTAGAATTGAAGTTTACTGATGGTGCATTAAAAGCA 1053

Qy     779 LeuAlaArgMetArgPheSerGluValValGluLysValAspAlaAlaGluAlaValArg 798
             :::|||   :::               :::::::::         |||         :::|||
Db    1054 ATTGCTGATTTA-----------GCTATCCAAAGAAATATGGGAGCTCGTGGTTTAAGA 1101

Qy     799 -----------------LeuLeuAspValAlaLeuGln---GlnSerAlaThrAspHis 811
                            ::::::|||:::      :::         |||      |||
Db    1102 ACAATCATTGAAAATTCAATTATGGATATTATGTATGAAACCCCAAGTGAAGAAGATATT 1161

Qy     812 AlaThrGlyThrIleAspMetAspLeuIleThrThrGlyValSerAlaSerGluArgIle 831
                :::      :::      |||:::||||||               |||      ||||||
Db    1162 GAATCAGTAGAAGTGACTAAAGACGTAATAACCCGT------CACGCACAGCCTCGCATT 1215

Qy     832 ArgArgAlaAsn 835
                |||   |||
Db    1216 ACTCGTAAAAAT 1227

Search completed: December  6, 2005, 14:25:28
Job time : 258 secs
```

```
OM protein - nucleic search, using frame_plus_p2n model

Run on:          December  6, 2005, 09:16:54 ; Search time 5712 Seconds
                                              (without alignments)
                                              7380.105 Million cell updates/sec

Title:           US-10-768-511-8
Perfect score:   4609
Sequence:        1 MENNDALDIGAVSSPYPSQS..........LGSLQGEGFLTVHGDIVKRV 901

Scoring table:   BLOSUM62
                 Xgapop 10.0 , Xgapext  0.5
                 Ygapop 10.0 , Ygapext  0.5
                 Fgapop  6.0 , Fgapext  7.0
                 Delop   6.0 , Delext   7.0

Searched:        41078325 seqs, 23393541228 residues

Total number of hits satisfying chosen parameters:      82156650

Minimum DB seq length: 0
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%
                 Maximum Match 100%
                 Listing first 45 summaries

Command line parameters:
-MODEL=frame+_p2n.model -DEV=xlh
-Q=/cgn2_1/USPTO_spool/US10768511/runat_02122005_102334_18419/app_query.fasta_1.1095
-DB=EST -QFMT=fastap -SUFFIX=rst -MINMATCH=0.1 -LOOPCL=0 -LOOPEXT=0
-UNITS=bits -START=1 -END=-1 -MATRIX=blosum62 -TRANS=human40.cdi -LIST=45
-DOCALIGN=200 -THR_SCORE=pct -THR_MAX=100 -THR_MIN=0 -ALIGN=15 -MODE=LOCAL
-OUTFMT=pto -NORM=ext -HEAPSIZE=500 -MINLEN=0 -MAXLEN=2000000000
-USER=US10768511_@CGN_1_1_4269_@runat_02122005_102334_18419 -NCPU=6 -ICPU=3
-NO_MMAP -LARGEQUERY -NEG_SCORES=0 -WAIT -DSPBLOCK=100 -LONGLOG
-DEV_TIMEOUT=120 -WARN_TIMEOUT=30 -THREADS=1 -XGAPOP=10 -XGAPEXT=0.5 -FGAPOP=6
-FGAPEXT=7 -YGAPOP=10 -YGAPEXT=0.5 -DELOP=6 -DELEXT=7

Database :       EST:*
                 1:  gb_est1:*
                 2:  gb_est2:*
                 3:  gb_est3:*
                 4:  gb_htc:*
                 5:  gb_est4:*
                 6:  gb_est5:*
                 7:  gb_est6:*
                 8:  gb_est7:*
                 9:  gb_gss1:*
                 10:  gb_gss2:*
                 11:  gb_gss3:*
```

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

	Result No.	Score	% Query Match	Length	DB	ID	Description
	1	2474	53.7	2699	4	CNS0A83A	BX820670 Arabidops
	2	2337.5	50.7	3168	10	CL959068	CL959068 OsIFCC002
	3	1822.5	39.5	3301	4	AK088796	AK088796 Mus muscu
	4	1818.5	39.5	3285	4	AK089999	AK089999 Mus muscu
	5	1809.5	39.3	3545	4	AK036653	AK036653 Mus muscu
	6	1808.5	39.2	3316	4	AK011743	AK011743 Mus muscu
	7	1803.5	39.1	2522	10	AY408033	AY408033 Homo sapi
	8	1799.5	39.0	2519	10	AY408035	AY408035 Mus muscu
	9	1661	36.0	2357	10	AY408034	AY408034 Pan trogl
	10	1334	28.9	1607	4	CR598188	CR598188 full-leng
	11	1300	28.2	3389	4	DQ057540	DQ057540 Antonospo
	12	1252	27.2	1833	4	CR606514	CR606514 full-leng
c	13	1191	25.8	879	8	DR495517	DR495517 WS0298.BR
	14	1165	25.3	881	7	CO123277	CO123277 GR__Eb05F
	15	1127	24.5	1156	7	CK210237	CK210237 FGAS02203
	16	1097	23.8	907	8	DR503163	DR503163 WS0298.B2
	17	1080	23.4	770	2	BF051102	BF051102 EST436277
	18	1060	23.0	788	3	BM412106	BM412106 EST586433
	19	1039	22.5	775	3	BM410852	BM410852 EST585179
c	20	1037	22.5	1082	5	BX836107	BX836107 BX836107
	21	1024	22.2	709	8	DR917805	DR917805 EST110934
	22	1020	22.1	750	6	CA258367	CA258367 SCCCRT300
	23	1004	21.8	755	8	DN232047	DN232047 MEST1067_
	24	997.5	21.6	1133	5	BX363316	BX363316 BX363316
	25	985	21.4	978	7	CK179662	CK179662 EST768982
	26	978.5	21.2	1109	7	CK027527	CK027527 AGENCOURT
	27	975.5	21.2	798	6	CF449139	CF449139 EST685484
	28	973.5	21.1	765	6	CF442356	CF442356 EST678701
	29	966	21.0	2387	4	BC023232	BC023232 Mus muscu
c	30	964.5	20.9	986	11	CNS06P5P	AL408947 T7 end of
	31	962	20.9	943	8	DN561478	DN561478 88797342
	32	954.5	20.7	727	7	CK072837	CK072837 68036rsic
	33	951.5	20.6	1142	3	BM479183	BM479183 AGENCOURT
	34	947.5	20.6	3295	4	AK037751	AK037751 Mus muscu
	35	943.5	20.5	2709	10	AY418605	AY418605 Homo sapi
	36	940.5	20.4	3283	4	AK088156	AK088156 Mus muscu
	37	939	20.4	2205	10	AY420810	AY420810 Homo sapi
	38	939	20.4	2452	4	CR591959	CR591959 full-leng
	39	939	20.4	2469	4	CR611599	CR611599 full-leng
	40	939	20.4	2508	4	CR615349	CR615349 full-leng
	41	936	20.3	2709	10	AY418607	AY418607 Mus muscu
	42	935	20.3	725	6	CA258377	CA258377 SCCCRT300
	43	933.5	20.3	1825	4	CR605715	CR605715 full-leng
	44	929.5	20.2	3177	10	CL980638	CL980638 OsIFCC035
	45	927.5	20.1	2480	4	CR601762	CR601762 full-leng

ALIGNMENTS

```
RESULT 1
CNS0A83A
LOCUS       CNS0A83A                2699 bp    mRNA    linear   HTC 06-FEB-2004
DEFINITION  Arabidopsis thaliana Full-length cDNA Complete sequence from clone
            GSLTPGH58ZF05 of Hormone Treated Callus of strain col-0 of
            Arabidopsis thaliana (thale cress).
ACCESSION   BX820670
VERSION     BX820670.1  GI:42467030
KEYWORDS    HTC; GSLT_cDNA.
SOURCE      Arabidopsis thaliana (thale cress)
  ORGANISM  Arabidopsis thaliana
            Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
            Spermatophyta; Magnoliophyta; eudicotyledons; core eudicotyledons;
            rosids; eurosids II; Brassicales; Brassicaceae; Arabidopsis.
REFERENCE   1  (bases 1 to 2699)
  AUTHORS   Castelli,V., Aury,J.M., Jaillon,O., Wincker,P., Clepet,C.,
            Menard,M., Cruaud,C., Quetier,F., Scarpelli,C., Schachter,V.,
            Temple,G., Caboche,M., Weissenbach,J. and Salanoubat,M.
  TITLE     Whole Genome Sequence Comparisons and 'Full-Length' cDNA Sequences:
            A Combined Approach to Evaluate and Improve Arabidopsis Genome
            Annotation
  JOURNAL   Unpublished
REFERENCE   2  (bases 1 to 2699)
  AUTHORS   Genoscope.
  TITLE     Direct Submission
  JOURNAL   Submitted (18-NOV-2003) Genoscope - Centre National de Sequencage :
            BP 191 91006 EVRY cedex - FRANCE (E-mail : seqref@genoscope.cns.fr
            - Web : www.genoscope.cns.fr)
COMMENT     The sequences are based on single pass reads.
            Life Technologies (a division of Invitrogen) members carried out
            full-length librairies construction : Temple G.
```

```
          Genoscope members carried out sequencing and annotation : Castelli
          V., Aury J.M., Jaillon O., Wincker P., Menard M., Cruaud C.,
          Schachter V., Weissenbach J., Salanoubat M.
          URGV INRA : Clepet C., Caboche M.
          Annotation is based on the June 2003 version of the Arabidopsis
          genome released by MIPS (Munich Information center for Protein
          Sequences). 5 prime and 3 prime are assembled with Phrap.
          http://www.genoscope.cns.fr/externe/sequences/Banque_Projet_EF/Full
          _length
          http://www.genoscope.cns.fr/cgi-bin/ggb/ggb?source=Arabidopsis.
FEATURES             Location/Qualifiers
     source          1..2699
                     /organism="Arabidopsis thaliana"
                     /mol_type="mRNA"
                     /db_xref="taxon:3702"
                     /clone="GSLTPGH58ZF05"
                     /tissue_type="Hormone Treated Callus"
                     /ecotype="Col-0"
                     /plasmid="pCMVSPORT_6"
     gene            complement(1..2699)
                     /gene="At2g16440"
ORIGIN

Alignment Scores:
Pred. No.:              1.21e-202     Length:         2699
Score:                  2474.00       Matches:        498
Percent Similarity:     70.92%        Conservative:   141
Best Local Similarity:  55.27%        Mismatches:     194
Query Match:            53.68%        Indels:         68
DB:                     4             Gaps:           13

US-10-768-511-8 (1-901) x CNS0A83A (1-2699)

Qy          9 IleGlyAlaValSerSerProTyrProSerGlnSerGluGlyValSerThrProLeuPro 28
Db         22 CTCGGCAACACCAACGACGGTCCTCCTTCTCCCGGAGAAAACGTTTCAAGCCCTATCGAA 81

Qy         29 GlnValThrSerProSerPheAspAsnAlaAlaSerProValAlaGlyArgArgAlaVal 48
Db         82 AACACATACTCT--------------------TCCCCGCCGCACTTCACCGT------ 114

Qy         49 ArgGlnThrProThrSerAlaValArgArgArgGlyArgGluThrAspSerAlaArgArg 68
Db        115 ------------------------CGCCGGCGAGGAAGATCC------------------ 132

Qy         69 ArgArgSerArgSerArgSerLeuGlyAsnSerValTyrSerSerProTyrAspAlaGly 88
Db        133 ---------------------------------------TCTACTCCCACGCAATTCGCA 153

Qy         89 ThrProGlyThrProGlyThrProValAlaThrProValTyrAlaThrProValGlyThr 108
Db        154 ACTCCTCCACCACCGCCTTCACGTCTCGCATCATCC---AATTCCACTCCT--------- 201

Qy        109 ProMetGlyThrProSerPheHisArgGlyThrProGlnTyrLysGlnArgSerGluLeu 128
Db        202 CCAACATCTCGTCCTTCCGCCGCCAGATCT------------AAAGGAAGAAATGGACAC 249

Qy        129 GlySerGlnGlyLysProLeuHisArgArgArgArgSerGlnSerArgGluProGlyHis 148
Db        250 GGAGGTGGCGGT-----------------GGTGGCGGTGGTGGCGGAGACCCAGGGACG 291

Qy        149 ArgSerProSerArgGluProSerAlaAspGlyArgProSerGluSerAlaGluProAsp 168
Db        292 CCAATGTCCACTGACGAGCCACTGCCTTCTTCCGATGACGGTGAAGAAGACGGAGGCGAC 351

Qy        169 AspThrLeuGlyGlyGluTyrAlaTyrValTrpGlyThrAsnValAsnIleProAspVal 188
Db        352 GATACTACTCCG---------ACTTTCGTTTGGGGAACTAACATCAGTGTGCAAGATGTT 402

Qy        189 LeuArgAlaIleArgArgPheLeuHisAsnTyr---------ArgSerSerAlaHisAsp 205
Db        403 AAATCTGCGATTGAAATGTTTGTTAAGCATTTCAGAGAAGCTAGAGAGAATAGTGATGAT 462

Qy        206 Leu------AsnSerLysTyrIleGlnIleIleGluGluThrValGluArgGluGluAsp 223
Db        463 TTGTTTAGAGAAGGGAAGTATATGGTTTCGATAAGGAAGGTGATTGAGATTGAAGGAGAA 522

Qy        224 ThrLeuAsnIleAspMetSerAspIleTyrAspHisAspProAspLeuTyrAlaLysIle 243
Db        523 TGGATTGATGTTGATGCGTTTGATGTGTTTGATTATGATCCTGATTTGTATAACAAGATG 582

Qy        244 ValArgTyrProLeuAspIleIleProLeuLeuAspThrGluCysGlnGluValAlaThr 263
Db        583 GTTCGTTATCCTTTGGAAGTTCTCGCTATCTTCGATATTGTGCTGATGGATATTGTTTCC 642

Qy        264 SerLeuLeuProThrPheGluLysHisIleGluAlaArgProPheAsnLeuLysAlaSer 283
Db        643 ACGATTAATCGTTTGTTCGAGAAACATGTTCAAGTTAGGATTTTTAACCTCAGGACATCT 702

Qy        284 ValHisMetArgGluLeuAsnProSerAspIleAspLysLeuValSerValLysGlyMet 303
Db        703 ACTTCCATGAGAAATCTCAATCCATCTGATATCGAGAAGATGATCTCTTTGAAAGGTATG 762

Qy        304 ValIleArgCysSerSerIleIleProGluIleLysGlyAlaPhePheLysCysLeuVal 323
Db        763 ATAATTCGGAGTAGCTCTATCATTCCTGAAATAAGGGAAGCTGTGTTTAGATGCCTTGTT 822

Qy        324 CysGlyHisSerProProLeuValThrValValLysGlyArgValGluGluProThrArg 343
Db        823 TGTGGTTACTTCTCTGACCCCATCATCGTTGATAGAGGGAAAATAAGTGAGCCTCCTACT 882

Qy        344 CysGluLysProGluCysAlaAlaArgAsnAlaMetSerLeuIleHisAsnArgCysThr 363
Db        883 TGCTTGAAACAAGAGTGCATGACCAAGAACTCCATGACACTAGTGCACAACCGATGCAGG 942

Qy        364 PheAlaAsnLysGlnIleValArgLeuGlnGluThrProAspAlaIleProGluGlyGlu 383
Db        943 TTTGCTGATAAGCAGATTGTGAGGCTTCAGGAAACACCTGATGAGATTCCAGAAGGAGGA 1002

Qy        384 ThrProHisThrValSerMetCysLeuTyrAsnThrMetValAspAlaValLysProGly 403
Db       1003 ACACCTCACACAGTTAGCTTGTTGCTGCATGATAAGCTAGTTGATAATGGGAAGCCTGGT 1062

Qy        404 AspArgIleGluValThrGlyValPheLysAlaMetAlaValArgValGlyProAsnGln 423
Db       1063 GATAGAATTGAGGTCACCGGAATTTACAGAGCAATGACTGTCCGAGTGGGACCTGCTCAC 1122

Qy        424 ArgThrLeuArgAlaLeuTyrLysThrTyrIleAspCysValHisValLysLysSerAsp 443
Db       1123 AGGACTGTGAAATCTGTGTTTAAGACCTATATCGATTGCCTTCACATAAAGAAAGCAAGT 1182

Qy        444 ArgGlyArgLeuGlnThrGluAspProMetGluMetAspLysGluAsnAspMetTyrAla 463
Db       1183 AAGTTAAGAATGTCTGCTGAGGATCCTATGGATGTTGACAAT----------------- 1224

Qy        464 GlyTyrHisGluSerAspThrSerGluAlaAlaAsnGluAlaLysIleGlnLysLeuLys 483
Db       1225 AGTCTGCGTAGAGTTGATGAAGATGTCGAGTTAGATGAGGAGAAGCTGAGGAAGTTTCAG 1284

Qy        484 GluLeuSerLysLeuProGlyIleTyrAspArgLeuSerArgSerLeuAlaProSerIle 503
Db       1285 GAACTCTCTAAACAACCAGATATATATGAGAGGCTTAGTAGATCACTAGCACCAAACATC 1344

Qy        504 TrpGluLeuGluAspIleLysLysGlyLeuLeuCysGlnLeuPheGlyGlyLysAlaLys 523
Db       1345 TGGGAGTTGGATGATGTTAAAAAGGGTCTTCTATGCCAGCTTTTTGGAGGGAATGCTTTG 1404

Qy        524 LysIleProSerGlyAlaSerPheArgGlyAspIleAsnValLeuLeuValGlyAspPro 543
Db       1405 AACTTAGCATCTGGTGCTAATTTTCGTGGTGACATCAACATCTTACTTGTTGGTGACCCT 1464

Qy        544 GlyThrSerLysSerGlnLeuLeuGlnTyrValHisLysIleAlaProArgGlyIleTyr 563
Db       1465 GGTACGAGCAAGTCCCAGCTGCTCCAGTACATTCATAAGCTTTCACCCCGTGGCATTTAC 1524

Qy        564 ThrSerGlyArgGlySerSerAlaValGlyLeuThrAlaTyrValThrLysAspProGlu 583
```

```
                 ||||||||||||||||||||||||||||||||||||||||||||||||||      ||||||||||||
Db        1525 ACAAGTGGGCGAGGGAGCTCAGCTGTTGGTTTAACAGCTTATGTAGCTAAAGATCCTGAG 1584

Qy         584 ThrArgGluThrValLeuGluSerGlyAlaLeuValLeuSerAspArgGlyIleCysCys 603
               |||  ||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        1585 ACAGGAGAAACTGTTTTGGAGAGTGGAGCTCTTGTTCTTAGTGACCGAGGTATCTGCTGT 1644

Qy         604 IleAspGluPheAspLysMetSerAspAsnAlaArgSerMetLeuHisGluValMetGlu 623
               |||||||||||||||||||||||||||||:::||||||||||||||||||||||||||||
Db        1645 ATTGATGAATTTGACAAAATGTCTGACAGTGCGAGGAGCATGCTACACGAGGTGATGGAA 1704

Qy         624 GlnGlnThrValSerValAlaLysGlyGlyIleIleAlaSerLeuAsnAlaArgThrSer 643
               ||||||||||||||:::||||||   ||||||||||||||||||||||||||||||||||
Db        1705 CAGCAGACTGTTTCAATAGCAAAGGCTGGTATTATTGCATCTCTAAATGCCAGAACCTCT 1764

Qy         644 ValLeuAlaCysAlaAsnProSerGlySerArgTyrAsnAlaArgLeuSerValIleAsp 663
               ||||||||||||||||||||||||||||||||||||||||   ||||||||||||||:::
Db        1765 GTGTTGGCTTGTGCAAATCCTAGTGGCTCACGTTATAATCCCCGGCTTTCTGTTATTGAG 1824

Qy         664 AsnIleGlnLeuProProThrLeuLeuSerArgPheAspLeuIleTyrLeuMetLeuAsp 683
               ||||||   ||||||||||||||||||||||||||||||||||||||||||:::||||||
Db        1825 AATATTCACCTTCCTCCAACCTTGCTATCTAGATTCGATTTGATCTACTTGATTCTTGAC 1884

Qy         684 LysProAspGluGlnAsnAspArgArgLeuAlaArgHisLeuValAlaLeuHisTyrGlu 703
               ||||||||||||||||    |||||||||||||||:::||||||||||||:::|||
Db        1885 AAGCCCGATGAGCAGACTGACCGAAGGCTTGCAAAGCATATTGTGGCCCTTCACTTTGAG 1944

Qy         704 AsnTyrGluValSerLysGlnAspAlaLeuAspLeuGlnThrLeuThrAlaTyrIleThr 723
               |||   |||   ::::::::::::|||:::|||:::   ||||||||   |||:::::::
Db        1945 AACGCTGAGAGTGCTCAGGAGGAAGCTATAGATATTACTACACTGACAACTTATGTTAGC 2004

Qy         724 TyrAlaArgGlnHisValHisProThrLeuSerAspGluAlaAlaGluAspLeuIleAsn 743
               |||||||||:::::::::||||||   ||||||||||||||||||||||:::|||
Db        2005 TATGCCCGCAAGAACATTCATCCTAAACTGTCAGATGAAGCTGCAGAGGAGTTGACCCGA 2064

Qy         744 GlyTyrValGluMetArgGlnLysGlyAsnPheProGlySerSerLysLysValIleThr 763
               ||||||||||||:::|||:::   |||   |||   ||||||||||||||||||||||||
Db        2065 GGATATGTTGAGCTGAGAAAAGCAGGGAAGTTTGCTGGCAGTAGCAAAAAGGTCATAACA 2124

Qy         764 AlaThrProArgGlnLeuGluSerMetIleArgIleSerGluAlaLeuAlaArgMetArg 783
               |||||||||||||:::||||||:::||||||:::||||||||||||||||||||||||||
Db        2125 GCAACTCCTAGGCAGATTGAAAGTTTGATCAGACTTAGTGAGGCCTTGGCTCGGATGCGC 2184

Qy         784 PheSerGluValValGluLysValAspAlaAlaGluAlaValArgLeuLeuAspValAla 803
               |||||||||   |||||||||   |||      ||||||   |||||||||   ||||||
Db        2185 TTCTCCGAATGGGTTGAAAAGCATGACGTGGACGAGGCATTTCGACTTCTCAGAGTTGCA 2244

Qy         804 LeuGlnGlnSerAlaThrAspHisAlaThrGlyThrIleAspMetAspLeuIleThrThr 823
               :::|||||||||||||||||||||||||||||||||||||||||||||||||   |||
Db        2245 ATGCAGCAATCAGCAACGGATCACGCCACAGGGACCATTGACATGGACCTGATAAATACT 2304

Qy         824 GlyValSerAlaSerGluArgIleArgArgAlaAsnLeuLeuAlaAlaLeuArgGluLeu 843
               |||||||||||||||||||||:::||||||         :::::::::|||::::::
Db        2305 GGAGTGTCAGCAAGTGAACGAATGAGGCGGGATACCTTTGCGTCGTCAATTAGAGACATA 2364

Qy         844 IleAlaAspLysIleSerProGlySerSerSerGlyLeuLysThrSerGlnLeuLeuGlu 863
                   :::|||:::      |||   ||||||   ::::::   |||:::|||||||||
Db        2365 GCTCTGGAGAAGATGCAAATTGGAGGATCATCA---ATGCGCTTATCCGAGTTACTTGAA 2421

Qy         864 AspIleArgSerGlnSer---------SerValAspValSerLeuGlnAspIleLysAsn 880
               :::::::::                  :::   ::::::      |||   |||::::::
Db        2422 GAACTGAAGAAGCATGGAGGCAACATAAACACCGAAATTCATCTTCATGATGTTAGAAAA 2481

Qy         881 AlaLeuGlySerLeuGlnGlyGluGlyPheLeuThrValHisGlyAspIleValLysArg 900
               |||:::   :::|||      ||||||||||||         ||||||   :::||||||
Db        2482 GCAGTCGCGACGCTAGCGAGTGAAGGATTTCTGGTAGCTGAAGGTGATAGAATCAAGAGA 2541

Qy         901 Val 901
               |||
Db        2542 GTA 2544
```

RESULT 2
CL959068

```
LOCUS       CL959068                3168 bp    DNA     linear   GSS 21-SEP-2004
DEFINITION  OsIFCC002182 Oryza sativa Express Library Oryza sativa (indica
            cultivar-group) genomic, genomic survey sequence.
ACCESSION   CL959068
VERSION     CL959068.1  GI:52372893
KEYWORDS    GSS.
SOURCE      Oryza sativa (indica cultivar-group)
  ORGANISM  Oryza sativa (indica cultivar-group)
            Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
            Spermatophyta; Magnoliophyta; Liliopsida; Poales; Poaceae;
            Ehrhartoideae; Oryzeae; Oryza.
REFERENCE   1  (bases 1 to 3168)
  AUTHORS   Ma,L., Wangc,J., Chen,C., Liu,X., Su,N., Li,L., Wang,X., Cao,M.,
            Jiao,Y., Sun,N., Zhang,X., Bao,J., Sun,D., Zhao,H., Yuan,L.,
            Wong,G.K.S., Deng,X.W. and Wang,J.
  TITLE     An analysis of transcriptional regulation of the rice genome and
            its comparison to Arabidopsis
  JOURNAL   Unpublished (2004)
COMMENT     Contact: Chen Chen
            Department of Bioinformatic
            Beijing Institute of Genomics
            Chinese Academy of Sciences, Beijing 101300, China
            Tel: 86-10-80481559
            Fax: 86-10-80488676
            Email: chenchen@genomics.org.cn
            Rice genomic sequence.
            Class: exon-trapped.
FEATURES             Location/Qualifiers
     source          1..3168
                     /organism="Oryza sativa (indica cultivar-group)"
                     /mol_type="genomic DNA"
                     /db_xref="taxon:39946"
                     /clone_lib="Oryza sativa Express Library"
                     /note="Oryza sativa exon trapped genomic sequences "
ORIGIN

Alignment Scores:
Pred. No.:              9.8e-191       Length:         3168
Score:                  2337.50        Matches:        492
Percent Similarity:     67.36%         Conservative:   123
Best Local Similarity:  53.89%         Mismatches:     175
Query Match:            50.72%         Indels:         125
DB:                     10             Gaps:           14

US-10-768-511-8 (1-901) x CL959068 (1-3168)

Qy           7 LeuAspIleGlyAlaValSerSerProTyrProSerGlnSerGluGlyValSerThrPro 26
               |||         ||||||||||||||                    |||:::|||
Db         746 TTGGTTTTTGCAGCCGTGTCGTCGCCG------------GATGTGCGGCCGTCGAGCCCT 793

Qy          27 LeuProGlnValThrSerProSerPheAspAsnAlaAlaSerProValAlaGlyArgArg 46
               ||||||                        |||::::::   |||   :::||||||||
Db         794 CTCCCGGCGACC------------------AACTCCTCTCCGCCTCAGTCGGGGCGGCGC 835

Qy          47 AlaValArgGlnThrProThr---------SerAlaValArgArgArgGlyArgGluThr 63
                        |||                  |||   ||||||   |||   :::
Db         836 GGCGGGGTGGCGGCGCCGGCGTGGCTCCGCCAGCCCGTACC-CGTCGTCCCCGTCCCTCG 894

Qy          64 AspSerAlaArgArgArgArgSerArgSerArgSerLeuGlyAsnSerValTyrSerSer 83
                  :::||||||||||||:::|||:::   :::|||
Db         895 GCGGGTTCGAGACGCCGCCGCACCCGGGCCGCCGCACTC--------------------- 933

Qy          84 ProTyrAspAla---------------GlyThrProGlyThrProGlyThrProValAla 98
               |||   |||               |||:::   |||      |||      |||
Db         934 CCTCCGGCGGCGCGGCGGCGGCGGCGCGGCAGCAGCGGCAGAACTGGACGGGGGGTGCGG 993

Qy          99 ThrProValTyrAlaThrProValGlyThrProMetGlyThrProSerPheHisArgGly 118
                  |||                              |||||||||
Db         994 TTCCCG------------------------------CCGACCCCGTCCACCCCCATGTCC 1023

Qy         119 ThrProGlnTyrLysGlnArgSerGluLeuGlySerGlnGlyLysProLeuHisArgArg 138
               |||               ||||||   |||   :::|||   |||      ||||||
Db        1024 ACCGACGATGTCCCGCTCTCCTCGGAGGCCGGGGA-CGAGGACACCCCCGAGACCGACGG 1082
```

```
Qy     139 ArgArgSerGlnSerArgGluProGlyHisArgSerProSerArgGluProSer-AlaAs 158
Db    1083 CGGCGGCGGCGGCGGCGC--------------------------------GGGCGCCGA 1109

Qy     158 pGlyArgProSerGluSerAlaGluProAspAspThrLeuGlyGlyGluTyrAlaTyrVa 178
Db    1110 CGCCACGCCG----------------------------------------GTGTTCGT 1127

Qy     178 lTrpGlyThrAsnValAsnIleProAspValLeuArgAlaIleArgArgPheLeuHisAs 198
Db    1128 CTGGGGCACCAACATCAGCGTGCAGGACGTCAACGCCGCCATCCTCCGGTTCCTCCGCCA 1187

Qy     198 nTyrArgSerSerAlaHisAspLeu------------------------AsnSerLysTy 210
Db    1188 CTTCCGC---GACCCGCGCGACGCCGGCCGCGTCGACCCGGTCATGGACGAGGGCAAGTA 1244

Qy     210 rIleGlnIleIleGluGluThrValGluArgGlu---GluAspThrLeuAsnIleAspMe 229
Db    1245 CATGCGCGCCATCCACCGCATCCTCGAGCTCGAGGGCGGGGAGTCGCTCGACGTCAACGC 1304

Qy     229 tSerAspIleTyrAspHisAspProAspLeuTyrAlaLysIleValArgTyrProLeuAs 249
Db    1305 CCACGACGTGTTCGACCACGACCCGGACCTCTACGGCAAGATGGTCAGGTACCCGCTCGA 1364

Qy     249 pIleIleProLeuLeuAspThrGluCysGlnGluValAlaThrSerLeuLeuProThrPh 269
Db    1365 GGTGCTCGCCATCTTTGATATCGTGCTCATGGACCTCGTCGCGCGCATCGAGCCCCTCTT 1424

Qy     269 eGluLysHisIleGluAlaArgProPheAsnLeuLysAlaSerValHisMetArgGluLe 289
Db    1425 CGAGAAGCACATCCAGACCAGGATCTACAACCTCAAGTCCTCGGTTTGCTTGAGGAATCT 1484

Qy     289 uAsnProSerAspIleAspLysLeuValSerValLysGlyMetValIleArgCysSerSe 309
Db    1485 CAACCCTTCTGATATCGAGAAGATGGTGTCCATCAAGGGTATGATAATTCGGTGCAGCTC 1544

Qy     309 rIleIleProGluIleLysGlyAlaPhePheLysCysLeuValCysGlyHisSerProPr 329
Db    1545 GGTGATTCCAGAGCTCAAGGAGGCTGTGTTCCGCTGCTTGGTTTGCGGGTTCTACTCTGA 1604

Qy     329 oLeuValThrValValLysGlyArgValGluGluProThrArgCysGluLysProGluCy 349
Db    1605 GCCTGTAATGGTTGATCGAGGGAGGGTTACTGAGCCTCACATCTGTCAGAAAGAACAATG 1664

Qy     349 sAlaAlaArgAsnAlaMetSerLeuIleHisAsnArgCysThrPheAlaAsnLysGlnIl 369
Db    1665 TAAAGCTACAAACTCTATGACTCTTGTGCATAACCGATGCAGGTTTGCAGATAAGCAGAT 1724

Qy     369 eValArgLeuGlnGluThrProAspAlaIleProGluGlyGluThrProHisThrValSe 389
Db    1725 CATAAAGTTGCAGGAAACACCAGATGAGATACCAGAAGGTGGCACTCCACATACTGTTAG 1784

Qy     389 rMetCysLeuTyrAsnThrMetValAspAlaValLysProGlyAspArgIleGluValTh 409
Db    1785 TGTCTTGATGCACGATAAGCTCGTAGATGCTGGAAAGCCTGGAGACAGGGTTGAGATAAC 1844

Qy     409 rGlyValPheLysAlaMetAlaValArgValGlyProAsnGlnArgThrLeuArgAlaLe 429
Db    1845 TGGGATATACAGAGCCATGAGTATTAGAGTTGGCCCAACTCAGAGGACAGTGAAGTCGAT 1904

Qy     429 uTyrLysThrTyrIleAspCysValHisValLysLysSerAspArgGlyArgLeuGlnTh 449
Db    1905 TTTC--------------------------AAGAAGACAGACAAGTCTAGACTTCATGT 1937

Qy     449 rGluAspProMetGluMetAspLysGluAsnAspMetTyrAlaGlyTyrHisGluSerAs 469
Db    1938 TGAGGACTCCATGGAAACTGATAACCCCAAT-------------------GCTAACAAGAC 1979

Qy     469 pThrSerGluAlaAlaAsnGluAlaLysIleGlnLysLeuLysGluLeuSerLysLeuPr 489
Db    1980 AACTGAAGATGATTTTCTCAGAGATAAGGTTGAGAAATTAAAAGAGTTGTCAAAGTTGCC 2039

Qy     489 oGlyIleTyrAspArgLeuSerArgSerLeuAlaProSerIleTrpGluLeuGluAspIl 509
Db    2040 AGATATATATGACAGATTAACTAGGTCATTGGCTCCAAACATATGGGAGCTGGACGACGT 2099

Qy     509 eLysLysGlyLeuLeuCysGlnLeuPheGlyGlyLysAlaLysLysIleProSerGlyAl 529
Db    2100 TAAAAGGGGCCTCCTTTGCCAGCTTTTGGTGGAAATGCTTTGAGGCTTCCTTCTGGAGC 2159

Qy     529 aSerPheArgGlyAspIleAsnValLeuLeuValGlyAspProGlyThrSerLysSerGl 549
Db    2160 TAGTTTCCGAGGCGACATCAATATTTTGCTTGTTGGTGATCCTGGAACAAGTAAATCCCA 2219

Qy     549 nLeuLeuGlnTyrValHisLysIleAlaProArgGlyIleTyrThrSerGlyArgGlySe 569
Db    2220 GCTTCTCCAATACATGCACAAACTGTCTCCTCGTGGCATTTATACAAGTGGCAGAGGAAG 2279

Qy     569 rSerAlaValGlyLeuThrAlaTyrValThrLysAspProGluThrArgGluThrValLe 589
Db    2280 TTCAGCTGTTGGCCTTACTGCATACGTTACCAAGGATCCTGAAACTGGTGAAACTGTTCT 2339

Qy     589 uGluSerGlyAlaLeuValLeuSerAspArgGlyIleCysCysIleAspGluPheAspLy 609
Db    2340 TGAGAGTGGAGCGCTTGTTTTGAGTGACAAAGGTGTTTGTTGTATTGATGAATTTGATAA 2399

Qy     609 sMetSerAspAsnAlaArgSerMetLeuHisGluValMetGluGlnGlnThrValSerVa 629
Db    2400 GATGTCTGATAATGCTCGAAGCATGTTACATGAGGTGATGGAACAACAGACTGTCTCCAT 2459

Qy     629 lAlaLysGlyGlyIleIleAlaSerLeuAsnAlaArgThrSerValLeuAlaCysAlaAs 649
Db    2460 TGCCAAGGCTGGAATAATTGCATCTTTAAATGCCAGAACATCAGTTCTAGCATGTGCAAA 2519

Qy     649 nProSerGlySerArgTyrAsnAlaArgLeuSerValIleAspAsnIleGlnLeuProPr 669
Db    2520 TCCTACTGAATCACGTTATAATCCAAGGCTCTCTGTGATTGACAATATCCATCTTCCTCC 2579

Qy     669 oThrLeuLeuSerArgPheAspLeuIleTyrLeuMetLeuAspLysProAspGluGlnAs 689
Db    2580 AACACTGCTTTCTAGGTTTGACCTCATTTATCTGATATTGGACAAGGCAGATGAGCAAAC 2639

Qy     689 nAspArgArgLeuAlaArgHisLeuValAlaLeuHisTyrGluAsnTyrGluValSerLy 709
Db    2640 TGATAGACGCCTGGCTAAGCATATTGTTTCGTTGCATTTTGAGAATCCAAACATAGAGGA 2699

Qy     709 sGlnAspAlaLeuAspLeuGlnThrLeuThrAlaTyrIleThrTyrAlaArgGlnHisVa 729
Db    2700 GCTCGAGGTCTTGGATTTGCCAACACTTGTAGCCTACATAAGTTATGCAAGGAAGCATAT 2759

Qy     729 lHisProThrLeuSerAspGluAlaAlaGluAspLeuIleAsnGlyTyrValGluMetAr 749
Db    2760 ACAACCACAGTTATCTGATGAAGCTGCAGAAGAATTGACCCGTGGCTATGTTGAGATGAG 2819

Qy     749 gGlnLysGlyAsnPheProGlySerSerLysLysValIleThrAlaThrProArgGlnLe 769
Db    2820 GAAAAGAGGAAACAGCCCTGGTAGCAGAAAGAAGGTCATAACTGCGACAGCTCGACAAAT 2879

Qy     769 uGluSerMetIleArgIleSerGluAlaLeuAlaArgMetArgPheSerGluValValGl 789
Db    2880 TGAGAGCTTGATTCGGCTCAGTGAAGCACTGGCCCGAATGCGATTCTCTGAAGTGGTTGA 2939

Qy     789 uLysValAspAlaAlaGluAlaValArgLeuLeuAspValAlaLeuGlnGlnSerAlaTh 809
Db    2940 AGTACGAGATGTTGTAGAGGCCTTCAGGCTTCTCGAAGTTGCCATGCAGCAATCGGCAAC 2999

Qy     809 rAspHisAlaThrGlyThrIleAspMetAspLeuIleThrThrGlyValSerAlaSerGl 829
Db    3000 TGATCATGCCACTGCTTGG---------------------------------------- 3018

Qy     829 uArgIleArgArgAlaAsnLeuLeuAlaAlaLeuArgGluLeuIleAlaAspLysIleSe 849
Db    3018 ------------------------------------------------------------ 3018

Qy     849 rProGlySerSerSerGlyLeuLysThrSerGlnLeuLeuGluAspIleArgSerGlnSe 869
```

```
                              ::::::       :::|||||||||:::||||||    |||||
Db       3019 ----------AGGCCCTCAGTGCGAATGATTGAGTTGCTGGAAGAAATTAGGAAGCAGAG 3068

Qy        869 rSerValAspValSerLeuGlnAspIleLysAsnAlaLeuGlySerLeuGlnGlyGluGl 889
              ||||:::::|||   |||   |||::::::   |||||||||:::|||       |||||
Db       3069 CTCTATGGAAGTTCATCTGCATGATCTTCGCGGTGCTCTTGGCACTCTGATGACAGAGGG 3128

Qy        889 yPheLeuThrValHisGlyAspIleValLysArgVal 901
              |   :::   :::|||||||||   |||||||||||
Db       3129 TGCCGTAGTCATCCATGGAGACAGCGTCAAGAGGGTC 3165

RESULT 3
AK088796
LOCUS       AK088796                3301 bp    mRNA    linear   HTC 03-APR-2004
DEFINITION  Mus musculus 2 days neonate thymus thymic cells cDNA, RIKEN
            full-length enriched library, clone:E430026B21 product:mini
            chromosome maintenance deficient 4 homolog (S. cerevisiae), full
            insert sequence.
ACCESSION   AK088796
VERSION     AK088796.1  GI:26353895
KEYWORDS    HTC; CAP trapper.
SOURCE      Mus musculus (house mouse)
  ORGANISM  Mus musculus
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Euarchontoglires; Glires; Rodentia;
            Sciurognathi; Muridae; Murinae; Mus.
REFERENCE   1
  AUTHORS   Carninci,P. and Hayashizaki,Y.
  TITLE     High-efficiency full-length cDNA cloning
  JOURNAL   Meth. Enzymol. 303, 19-44 (1999)
   PUBMED   10349636
REFERENCE   2
  AUTHORS   Carninci,P., Shibata,Y., Hayatsu,N., Sugahara,Y., Shibata,K.,
            Itoh,M., Konno,H., Okazaki,Y., Muramatsu,M. and Hayashizaki,Y.
  TITLE     Normalization and subtraction of cap-trapper-selected cDNAs to
            prepare full-length cDNA libraries for rapid discovery of new genes
  JOURNAL   Genome Res. 10 (10), 1617-1630 (2000)
   PUBMED   11042159
REFERENCE   3
  AUTHORS   Shibata,K., Itoh,M., Aizawa,K., Nagaoka,S., Sasaki,N., Carninci,P.,
            Konno,H., Akiyama,J., Nishi,K., Kitsunai,T., Tashiro,H., Itoh,M.,
            Sumi,N., Ishii,Y., Nakamura,S., Hazama,M., Nishine,T., Harada,A.,
            Yamamoto,R., Matsumoto,H., Sakaguchi,S., Ikegami,T., Kashiwagi,K.,
            Fujiwake,S., Inoue,K., Togawa,Y., Izawa,M., Ohara,E., Watahiki,M.,
            Yoneda,Y., Ishikawa,T., Ozawa,K., Tanaka,T., Matsuura,S., Kawai,J.,
            Okazaki,Y., Muramatsu,M., Inoue,Y., Kira,A. and Hayashizaki,Y.
  TITLE     RIKEN integrated sequence analysis (RISA) system--384-format
            sequencing pipeline with 384 multicapillary sequencer
  JOURNAL   Genome Res. 10 (11), 1757-1771 (2000)
   PUBMED   11076861
REFERENCE   4
  AUTHORS   The RIKEN Genome Exploration Research Group Phase II Team and the
            FANTOM Consortium.
  TITLE     Functional annotation of a full-length mouse cDNA collection
  JOURNAL   Nature 409, 685-690 (2001)
REFERENCE   5
  AUTHORS   The FANTOM Consortium and the RIKEN Genome Exploration Research
            Group Phase I & II Team.
  TITLE     Analysis of the mouse transcriptome based on functional annotation
            of 60,770 full-length cDNAs
  JOURNAL   Nature 420, 563-573 (2002)
REFERENCE   6  (bases 1 to 3301)
  AUTHORS   Adachi,J., Aizawa,K., Akimura,T., Arakawa,T., Bono,H., Carninci,P.,
            Fukuda,S., Furuno,M., Hanagaki,T., Hara,A., Hashizume,W.,
            Hayashida,K., Hayatsu,N., Hiramoto,K., Hiraoka,T., Hirozane,T.,
            Hori,F., Imotani,K., Ishii,Y., Itoh,M., Kagawa,I., Kasukawa,T.,
            Katoh,H., Kawai,J., Kojima,Y., Kondo,S., Konno,H., Kouda,M.,
            Koya,S., Kurihara,C., Matsuyama,T., Miyazaki,A., Murata,M.,
            Nakamura,M., Nishi,K., Nomura,K., Numazaki,R., Ohno,M., Ohsato,N.,
            Okazaki,Y., Saito,R., Saitoh,H., Sakai,C., Sakai,K., Sakazume,N.,
            Sano,H., Sasaki,D., Shibata,K., Shinagawa,A., Shiraki,T.,
            Sogabe,Y., Tagami,M., Tagawa,A., Takahashi,F., Takaku-Akahira,S.,
            Takeda,Y., Tanaka,T., Tomaru,A., Toya,T., Yasunishi,A.,
            Muramatsu,M. and Hayashizaki,Y.
  TITLE     Direct Submission
  JOURNAL   Submitted (16-APR-2002) Yoshihide Hayashizaki, The Institute of
            Physical and Chemical Research (RIKEN), Laboratory for Genome
            Exploration Research Group, RIKEN Genomic Sciences Center (GSC),
            RIKEN Yokohama Institute; 1-7-22 Suehiro-cho, Tsurumi-ku, Yokohama,
            Kanagawa 230-0045, Japan (E-mail:genome-res@gsc.riken.jp,
            URL:http://genome.gsc.riken.jp/, Tel:81-45-503-9222,
            Fax:81-45-503-9216)
COMMENT     cDNA library was prepared and sequenced in Mouse Genome
            Encyclopedia Project of Genome Exploration Research Group in Riken
            Genomic Sciences Center and Genome Science Laboratory in RIKEN.
            Division of Experimental Animal Research in Riken contributed to
            prepare mouse tissues.
            Tissues were provided by Dr. John Todd (Dept. of Medical Genetics
            Wellcome Trust Centre for Molecular Mechanisms in Disease Wellcome
            Trust/MRC building Addenbrookes Hospital Cambridge) whose
            assistance we gratefully acknowledge.
            Please visit our web site for further details.
            URL:http://genome.gsc.riken.jp/
            URL:http://fantom.gsc.riken.jp/.
FEATURES             Location/Qualifiers
     source          1..3301
                     /organism="Mus musculus"
                     /mol_type="mRNA"
                     /strain="NOD"
                     /db_xref="FANTOM_DB:E430026B21"
                     /db_xref="taxon:10090"
                     /clone="E430026B21"
                     /cell_type="thymic cells"
                     /tissue_type="thymus"
                     /clone_lib="RIKEN full-length enriched mouse cDNA library"
                     /dev_stage="2 days neonate"
     CDS             141..2729
                     /note="unnamed protein product; mini chromosome
                     maintenance deficient 4 homolog (S. cerevisiae)
                     (MGD|MGI:103199, GB|NM_008565, evidence: BLASTN, 100%,
                     match=3221)
                     putative"
                     /codon_start=1
                     /protein_id="BAC40578.1"
                     /db_xref="GI:26353896"
                     /translation="MSSPASTPSRRSSRRGRVTPTQSLRSEESRSSPNRRRRGEDSST
                     GELLPMPTSPGADLQSPPAQNALFSSPPQMHSLAIPLDFDVSSPLTYGTPSSRVEGTP
                     RSGVRGTPVRQRPDLGSARKGLQVDLQSDGAAAEDIVPSEQSLGQKLVIWGTDVNVAT
                     CKENFQRFLQCFTDPLAKEEENVGIDITQPLYMQQLGEINITGEPFLNVNCEHIKSFS
                     KNLYRQLISYPQEVIPTFDMAVNEIFFDRYPDSILEHQIQVRPFNALKTKSMRNLNPE
                     DIDQLITISGMVIRTSQLIPEMQEAFFQCQVCAHTTRVEIDRGRIAEPCSCVHCHTTH
                     SMALIHNRSFFSDKQMIKLQESPEDMPAGQTPHTIVLFAHNDLVDKVQPGDRVNVTGI
                     YRAVPIRVNPRVSNVKSVYKTHIDVIHYRKTDAKRLHGLDEEAEQKLFSEKRVKLLKE
                     LSRKPDIYERLASALAPSIYEHEDIKKGILLQLFGGTRKDFSHTGRGKFRAEINILLC
                     GDPGTSKSQLLQYVYNLVPRGQYTSGKGSSAVGLTAYVMKDPETRQLVLQTGALVLSD
                     NGICCIDEFDKMNESTRSVLHEVMEQQTLSIAKAGIICQLNARTSVLAAANPIESQWN
                     PKKTTIENIQLPHTLLSRFDLIFLMLDPQDEAYDRRLAHHLVSLYYQSEEQVEEEFLD
                     MAVLKDYIAYAHSTIMPRLSEEASQALIEAYVNMRKIGSSRGMVSAYPRQLESLIRLA
                     EAHAKVRFSNKVEAIDVEEAKRLHREALKQSATDPRTGIVDISILTTGMSATSRKRKE
                     ELAEALRKLILSKGKTPALKYQQLFEDIRGQSDTAITKDMFEEALRALADDDFLTVTG
                     KTVRLL"
     polyA_signal    3284..3289
                     /note="putative"
     polyA_site      3301
                     /note="putative"
ORIGIN

Alignment Scores:
Pred. No.:              3.73e-146      Length:         3301
Score:                  1822.50        Matches:        403
Percent Similarity:     59.57%         Conservative:   151
Best Local Similarity:  43.33%         Mismatches:     267
Query Match:            39.54%         Indels:         109
DB:                     4              Gaps:           17

US-10-768-511-8 (1-901) x AK088796 (1-3301)
```

```
Qy     14 SerProTyrProSerGlnSerGluGlyValSerThrProLeuProGlnValThrSerPro 33
Db    126 AGCAACCACAGCACCATGTCGTCCCCGGCATCCACCCCG-------------------- 164

Qy     34 SerPheAspAsnAlaAlaSerProValAlaGlyArgArgAlaValArgGlnThrProThr 53
Db    165 ------------------AGCCGCCGCAGCAGCCGACGCGGA---CGAGTCACCCCAACC 203

Qy     54 SerAlaValArgArgArgGlyArgGluThrAspSerAlaArgArgArgArgSerArgSer 73
Db    204 CAGTCCCTTCGAAGTGAGGAAAGCAGGTCGTCACCCAATCGGAGACGTAGAGGCGAAGAT 263

Qy     74 ArgSerLeuGly------------------------------------------------- 77
Db    264 TCTTCCACCGGAGAGCTACTGCCAATGCCCACCTCACCAGGAGCCGACCTGCAGAGCCCA 323

Qy     78 ---------AsnSerValTyrSerSerPro------------------------------ 84
Db    324 CCTGCACAGAATGCCTTGTTTTCCAGCCCTCCTCAGATGCATTCTTTAGCTATTCCTTTG 383

Qy     85 ---TyrAspAlaGlyThrProGlyThrProGlyThrProValAlaThrProValTyrAla 103
Db    384 GACTTTGATGTTAGTTCACCATTGACA--------------------------TATGGC 416

Qy    104 ThrProValGlyThrProMetGlyThrPro---SerPheHisArgGlyThrProGlnTyr 122
Db    417 ACTCCCAGCTCGCGAGTGGAAGGAACCCCAAGAAGTGGGGTGAGAGGCACACCT---GTA 473

Qy    123 LysGlnArgSerGluLeuGlySerGlnGlyLysProLeuHisArgArgArgArgSerGln 142
Db    474 AGGCAGAGGCCAGATCTGGGCTCAGCACGAAAGGGTTTGCAGGTGGATCTGCAGTCT--- 530

Qy    143 SerArgGluProGlyHisArgSerProSerArgGluProSerAlaAspGlyArgProSer 162
Db    531 ---------------------------------------------GATGGCGCAGCAGCA 545

Qy    163 GluSerAlaGluProAspAspThrLeuGlyGlyGluTyrAlaTyrValTrpGlyThrAsn 182
Db    546 GAAGACATCGTACCAAGTGAACAGTCTCTAGGCCAAAAGCTTGTGATTTGGGGAACAGAT 605

Qy    183 ValAsnIleProAspValLeuArgAlaIleArgArgPheLeuHisAsnTyr--------- 199
Db    606 GTGAATGTGGCAACATGTAAAGAGAATTTTCAGAGATTCCTTCAGTGTTTTACTGATCCT 665

Qy    200 ---------------ArgSerSerAlaHisAspLeuAsnSerLys---TyrIleGlnIle 213
Db    666 CTGGCCAAAGAAGAAGAAAATGTTGGCATAGATATTACTCAACCTTTGTACATGCAACAA 725

Qy    214 IleGluGluThrValGluArgGluGluAspThrLeuAsnIleAspMetSerAspIleTyr 233
Db    726 CTTGGAGAGATTAATATTACAGGAGAGCCATTTTTAAATGTGAACTGCGAACACATAAAA 785

Qy    234 AspHisAspProAspLeuTyrAlaLysIleValArgTyrProLeuAspIleIleProLeu 253
Db    786 TCATTTAGCAAAAATCTGTATAGACAGCTCATCTCCTACCCACAGGAGGTTATACCAACC 845

Qy    254 LeuAspThrGluCysGlnGluValAlaThrSerLeuLeuPro------ThrPheGluLys 271
Db    846 TTTGACATGGCTGTCAATGAGATCTTCTTTGACCGTTATCCTGACTCCATCTTAGAACAT 905

Qy    272 HisIleGluAlaArgProPheAsnLeuLysAlaSerValHisMetArgGluLeuAsnPro 291
Db    906 CAGATTCAAGTCAGACCTTTTAATGCGTTGAAGACAAAGAGTATGAGAAACTTGAATCCA 965

Qy    292 SerAspIleAspLysLeuValSerValLysGlyMetValIleArgCysSerSerIleIle 311
Db    966 GAAGACATTGATCAGCTCATCACCATCAGTGGCATGGTCATCAGAACATCACAGCTGATT 1025

Qy    312 ProGluIleLysGlyAlaPhePheLysCysLeuValCysGlyHisSerProProLeuVal 331
Db   1026 CCGGAGATGCAGGAGGCCTTTTTCCAATGCCAAGTCTGTGCCCACACC---ACCCGGGTG 1082

Qy    332 ThrValValLysGlyArgValGluGluProThrArgCysGluLysProGluCysAlaAla 351
Db   1083 GAGATAGATCGAGGCAGAATTGCTGAGCCCTGCAGTTGT------GTGCACTGCCACACT 1136

Qy    352 ArgAsnAlaMetSerLeuIleHisAsnArgCysThrPheAlaAsnLysGlnIleValArg 371
Db   1137 ACCCACAGCATGGCACTGATCCACAACCGATCATTCTTCTCTGACAAGCAAATGATCAAA 1196

Qy    372 LeuGlnGluThrProAspAlaIleProGluGlyGluThrProHisThrValSerMetCys 391
Db   1197 CTTCAAGAGTCTCCTGAAGACATGCCTGCTGGGCAGACACCTCACACTATTGTCCTTTTT 1256

Qy    392 LeuTyrAsnThrMetValAspAlaValLysProGlyAspArgIleGluValThrGlyVal 411
Db   1257 GCCCACAATGACCTTGTTGACAAGGTTCAACCAGGGGACAGAGTGAACGTCACAGGCATA 1316

Qy    412 PheLysAlaMetAlaValArgValGlyProAsnGlnArgThrLeuArgAlaLeuTyrLys 431
Db   1317 TATCGAGCAGTACCAATTCGAGTTAATCCAAGAGTGAGCAACGTGAAGTCTGTCTATAAA 1376

Qy    432 ThrTyrIleAspCysValHisValLysLysSerAspArgGlyArgLeuGlnThrGluAsp 451
Db   1377 ACCCACATTGATGTCATTCATTATCGGAAAACGGATGCAAAACGTCTGCAT--------- 1427

Qy    452 ProMetGluMetAspLysGluAsnAspMetTyrAlaGlyTyrHisGluSerAspThrSer 471
Db   1428 ------------------------------------GGCCTTGATGAAGAAGCAGAACAG 1451

Qy    472 GluAlaAlaAsnGluAlaLysIleGlnLysLeuLysGluLeuSerLysLeuProGlyIle 491
Db   1452 AAACTTTTTCAGAGAAACGTGTGAAATTGCTTAAGGAACTTTCCAGGAAGCCAGATATT 1511

Qy    492 TyrAspArgLeuSerArgSerLeuAlaProSerIleTrpGluLeuGluAspIleLysLys 511
Db   1512 TATGAGCGGCTTGCTTCAGCCTTGGCTCCCAGCATTTATGAACATGAAGATATCAAAAAG 1571

Qy    512 GlyLeuLeuCysGlnLeuPheGlyGlyLysAlaLysLysIle------ProSerGlyAla 529
Db   1572 GGAATCTTACTTCAGCTCTTTGGTGGAACAAGGAAGGATTTCAGTCACACTGGGAGGGGT 1631

Qy    530 SerPheArgGlyAspIleAsnValLeuLeuValGlyAspProGlyThrSerLysSerGln 549
Db   1632 AAATTCCGTGCTGAGATCAACATCCTTCTGTGTGGGGACCCTGGCACCAGCAAGTCCCAG 1691

Qy    550 LeuLeuGlnTyrValHisLysIleAlaProArgGlyIleTyrThrSerGlyArgGlySer 569
Db   1692 CTGCTACAGTATGTGTACAACCTGGTCCCCAGAGGCCAGTACACGTCTGGAAAAGGCTCC 1751

Qy    570 SerAlaValGlyLeuThrAlaTyrValThrLysAspProGluThrArgGluThrValLeu 589
Db   1752 AGTGCGGTCGGCCTCACCGCCTATGTGATGAAAGACCCTGAGACCAGGCAGCTTGTCCTC 1811

Qy    590 GluSerGlyAlaLeuValLeuSerAspArgGlyIleCysCysIleAspGluPheAspLys 609
Db   1812 CAGACAGGTGCCCTCGTCCTGAGTGACAATGGGATATGCTGCATCGATGAGTTTGACAAA 1871

Qy    610 MetSerAspAsnAlaArgSerMetLeuHisGluValMetGluGlnGlnThrValSerVal 629
Db   1872 ATGAATGAAAGCACAAGGTCTGTGCTGCATGAGGTCATGGAACAGCAGACTCTGTCCATT 1931

Qy    630 AlaLysGlyGlyIleIleAlaSerLeuAsnAlaArgThrSerValLeuAlaCysAlaAsn 649
Db   1932 GCAAAGGCTGGGATCATCTGTCAGCTCAATGCGCGCACCTCTGTCCTGGCAGCAGCAAAT 1991

Qy    650 ProSerGlySerArgTyrAsnAlaArgLeuSerValIleAspAsnIleGlnLeuProPro 669
Db   1992 CCTATTGAGTCTCAGTGGAATCCTAAAAAAACAACCATTGAAAATATCCAACTACCGCAC 2051

Qy    670 ThrLeuLeuSerArgPheAspLeuIleTyrLeuMetLeuAspLysProAspGluGlnAsn 689
Db   2052 ACATTGTTGTCAAGGTTTGATCTCATTTTCCTCATGCTAGACCCTCAGGATGAGGCATAT 2111

Qy    690 AspArgArgLeuAlaArgHisLeuValAlaLeuHisTyrGluAsnTyrGluValSerLys 709
```

```
                 ||||||||||||||     |||||||||:::|||:::|||::::::    |||        :::
Db       2112 GACCGGCGTCTAGCTCATCACCTGGTTTCATTGTACTACCAAAGTGAGGAGCAAGTGGAG 2171

Qy        710 GlnAspAlaLeuAspLeuGlnThrLeuThrAlaTyrIleThrTyrAlaArgGlnHisVal 729
              ::::::   |||||||:::          |||       ||||||    ||||||          :::
Db       2172 GAGGAGTTCCTGGACATGGCCGTGCTGAAAGACTACATTGCATATGCCCATAGTACCATC 2231

Qy        730 HisProThrLeuSerAspGluAlaAlaGluAspLeuIleAsnGlyTyrValGluMetArg 749
                 |||   ||||||:::||||||::::::        ||||||       ||||||    ||||||
Db       2232 ATGCCCCGACTGAGTGAGGAGGCCAGCCAGGCTCTCATTGAGGCTTATGTAAACATGAGG 2291

Qy        750 GlnLysGlyAsnPheProGlySerSerLysLysValIleThrAlaThrProArgGlnLeu 769
              :::                 |||||||||:::   :::::::::|||    |||||||||||
Db       2292 AAG------------ATTGGGAGTAGCCGGGGGATGGTTTCTGCTTACCCTCGACAGCTA 2339

Qy        770 GluSerMetIleArgIleSerGluAlaLeuAlaArgMetArgPheSerGluValValGlu 789
              ||||||:::||||||::::::||||||       |||:::::::||||||||        ||||||
Db       2340 GAGTCATTAATTCGCTTAGCAGAAGCCCATGCTAAAGTAAGATTTTCAAACAAAGTTGAA 2399

Qy        790 LysValAspAlaAlaGluAlaValArgLeuLeuAspValAlaLeuGlnGlnSerAlaThr 809
                 :::|||            ||||||   ||||||            ||||||:::||||||||||||
Db       2400 GCAATTGATGTGGAAGAGGCAAAACGCCTCCACCGGGAGGCTCTGAAGCAGTCTGCAACT 2459

Qy        810 AspHisAlaThrGlyThrIleAspMetAspLeuIleThrThrGlyValSerAlaSerGlu 829
              |||          ||||||      :::|||:::      ::::::||||||||:::||||||:::
Db       2460 GACCCTCGTACTGGCATTGTGGATATTTCTATTCTTACTACAGGAATGAGTGCCACTTCT 2519

Qy        830 ArgIleArgArgAlaAsnLeuLeuAlaAlaLeuArgGluLeuIleAlaAspLysIleSer 849
              |||    |||:::         |||       |||||||||:::||||||                :::|||
Db       2520 CGTAAACGGAAAGAAGAATTAGCTGAAGCATTGAGAAAACTTATT---------TTATCT 2570

Qy        850 ProGlySerSerSerGlyLeuLysThrSerGlnLeuLeuGluAspIleArgSerGlnSer 869
                 |||    :::      ||||||        ||||||    |||||||||||||    ||||||
Db       2571 AAGGGTAAAACACCAGCCTTAAAGTACCAACAGCTGTTTGAGGATATTCGGGGACAGTCT 2630

Qy        870 SerValAspValSerLeuGlnAspIleLysAsnAlaLeuGlySerLeuGlnGlyGluGly 889
                       ::::::            :::    ||||||    :::|||      :::
Db       2631 GACACAGCAATTACCAAGGACATGTTTGAAGAAGCCCTGCGAGCTTTGGCTGATGATGAT 2690

Qy        890 PheLeuThrValHisGlyAspIleValLys 899
              ||||||||||||     |||          |||:::
Db       2691 TTCCTAACAGTGACTGGGAAGACTGTCCGC 2720
```

RESULT 4
AK089999
LOCUS   AK089999   3285 bp   mRNA   linear   HTC 03-APR-2004
DEFINITION Mus musculus brain CRL-1443 BC3H1 cDNA, RIKEN full-length enriched library, clone:G430050N09 product:mini chromosome maintenance deficient 4 homolog (S. cerevisiae), full insert sequence.
ACCESSION   AK089999
VERSION   AK089999.1 GI:26354818
KEYWORDS   HTC; CAP trapper.
SOURCE   Mus musculus (house mouse)
  ORGANISM Mus musculus
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Euarchontoglires; Glires; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
REFERENCE   1
  AUTHORS   Carninci,P. and Hayashizaki,Y.
  TITLE   High-efficiency full-length cDNA cloning
  JOURNAL   Meth. Enzymol. 303, 19-44 (1999)
   PUBMED   10349636
REFERENCE   2
  AUTHORS   Carninci,P., Shibata,Y., Hayatsu,N., Sugahara,Y., Shibata,K., Itoh,M., Konno,H., Okazaki,Y., Muramatsu,M. and Hayashizaki,Y.
  TITLE   Normalization and subtraction of cap-trapper-selected cDNAs to prepare full-length cDNA libraries for rapid discovery of new genes
  JOURNAL   Genome Res. 10 (10), 1617-1630 (2000)
   PUBMED   11042159
REFERENCE   3
  AUTHORS   Shibata,K., Itoh,M., Aizawa,K., Nagaoka,S., Sasaki,N., Carninci,P., Konno,H., Akiyama,J., Nishi,K., Kitsunai,T., Tashiro,H., Itoh,M., Sumi,N., Ishii,Y., Nakamura,S., Hazama,M., Nishine,T., Harada,A., Yamamoto,R., Matsumoto,H., Sakaguchi,S., Ikegami,T., Kashiwagi,K., Fujiwake,S., Inoue,K., Togawa,Y., Izawa,M., Ohara,E., Watahiki,M., Yoneda,Y., Ishikawa,T., Ozawa,K., Tanaka,T., Matsuura,S., Kawai,J., Okazaki,Y., Muramatsu,M., Inoue,Y., Kira,A. and Hayashizaki,Y.
  TITLE   RIKEN integrated sequence analysis (RISA) system--384-format sequencing pipeline with 384 multicapillary sequencer
  JOURNAL   Genome Res. 10 (11), 1757-1771 (2000)
   PUBMED   11076861
REFERENCE   4
  AUTHORS   The RIKEN Genome Exploration Research Group Phase II Team and the FANTOM Consortium.
  TITLE   Functional annotation of a full-length mouse cDNA collection
  JOURNAL   Nature 409, 685-690 (2001)
REFERENCE   5
  AUTHORS   The FANTOM Consortium and the RIKEN Genome Exploration Research Group Phase I & II Team.
  TITLE   Analysis of the mouse transcriptome based on functional annotation of 60,770 full-length cDNAs
  JOURNAL   Nature 420, 563-573 (2002)
REFERENCE   6 (bases 1 to 3285)
  AUTHORS   Adachi,J., Aizawa,K., Akimura,T., Arakawa,T., Bono,H., Carninci,P., Fukuda,S., Furuno,M., Hanagaki,T., Hara,A., Hashizume,W., Hayashida,K., Hayatsu,N., Hiramoto,K., Hiraoka,T., Hirozane,T., Hori,F., Imotani,K., Ishii,Y., Itoh,M., Kagawa,I., Kasukawa,T., Katoh,H., Kawai,J., Kojima,Y., Kondo,S., Konno,H., Kouda,M., Koya,S., Kurihara,C., Matsuyama,T., Miyazaki,A., Murata,M., Nakamura,M., Nishi,K., Nomura,K., Numazaki,R., Ohno,M., Ohsato,N., Okazaki,Y., Saito,R., Saitoh,H., Sakai,C., Sakai,K., Sakazume,N., Sano,H., Sasaki,D., Shibata,K., Shinagawa,A., Shiraki,T., Sogabe,Y., Tagami,M., Tagawa,A., Takahashi,F., Takaku-Akahira,S., Takeda,Y., Tanaka,T., Tomaru,A., Toya,T., Yasunishi,A., Muramatsu,M. and Hayashizaki,Y.
  TITLE   Direct Submission
  JOURNAL   Submitted (16-APR-2002) Yoshihide Hayashizaki, The Institute of Physical and Chemical Research (RIKEN), Laboratory for Genome Exploration Research Group, RIKEN Genomic Sciences Center (GSC), RIKEN Yokohama Institute; 1-7-22 Suehiro-cho, Tsurumi-ku, Yokohama, Kanagawa 230-0045, Japan (E-mail:genome-res@gsc.riken.jp, URL:http://genome.gsc.riken.jp/, Tel:81-45-503-9222, Fax:81-45-503-9216)
COMMENT   cDNA library was prepared and sequenced in Mouse Genome Encyclopedia Project of Genome Exploration Research Group in Riken Genomic Sciences Center and Genome Science Laboratory in RIKEN. Division of Experimental Animal Research in Riken contributed to prepare mouse tissues.
            Please visit our web site for further details.
            URL:http://genome.gsc.riken.jp/
            URL:http://fantom.gsc.riken.jp/.
FEATURES   Location/Qualifiers
     source   1. .3285
                     /organism="Mus musculus"
                     /mol_type="mRNA"
                     /strain="C3H"
                     /db_xref="FANTOM_DB:G430050N09"
                     /db_xref="taxon:10090"
                     /clone="G430050N09"
                     /cell_line="CRL-1443 BC3H1"
                     /tissue_type="brain"
                     /clone_lib="RIKEN full-length enriched mouse cDNA library"
     CDS   125. .2713
                     /note="unnamed protein product; mini chromosome maintenance deficient 4 homolog (S. cerevisiae) (MGD|MGI:103199, GB|NM_008565, evidence: BLASTN, 100%, match=3221)
                     putative"
                     /codon_start=1
                     /protein_id="BAC41036.1"
                     /db_xref="GI:26354819"
                     /translation="MSSPASTPSRRSSRRGRVTPTQSLRSEESRSSPNRRRRGEDSST
                     GELLPMPTSPGADLQSPPAQNALFSSPPQMHSLAIPLDFDVSSPLTYGTPSSRVEGTP
                     RSGVRGTPVRQRPDLGSARKGLQVDLQSDGAAAEDIVPSEQSLGQKLVIWGTDVNVAT
                     CKENFQRFLQCFTDPLAQEEENVGIDITQPLYMQQLGEINITGEPFLNVNCEHIKSFS
                     KNLYRQLISYPQEVIPTFDMAVNEIFFDRYPDSILEHQIQVRPFNALKTKSMRNLNPE

```
                    DIDQLITISGMVIRTSQLIPEMQEAFFQCQVCAHTTRVEIDRGRIAEPCSCVHCHTTH
                    SMALIHNRSFFSDKQMIKLQESPEDMPAGQTPHTIVLFAHNDLVDKVQPGDRVNVTGI
                    YRAVPIRVNPRVSNVKSVYKTHIDVIHYRKTDAKRLHGLDEEAEQKLFSEKRVKLLKE
                    LSRKPDIYERLASALAPSIYEHEDIKKGILLQLFGGTRKDFSHTGRGKFRAEINILLC
                    GDPGTSKSQLLQYVYNLVPRGQYTSGKGSSAVGLTAYVMKDPETRQLVLQTGALVLSD
                    NGICCIDKFDKMNESTRSVLHEVMEQQTLSIAKAGIICQLNARTSVLAAANPIESQWN
                    PKKTTIENIQLPHTLLSRFDLIFLMLDPQDEAYDRRLAHHLVSLYYQSEEQVEEEFLD
                    MAVLKDYIAYAHSTIMPRLSEEASQALIEAYVNMRKIGSSRGMVSAYPRQLESLIRLA
                    EAHAKVRFSNKVEAIDVEEAKRLHREALKQSATDPRTGMVDISILTTGMSATSRKRKE
                    ELAEALRKLILSKGKTPALKYQQLFEDIRGQSDTAITKDMFEEALRALADDDFLTVTG
                    KTVRLL"
     polyA_signal   3268..3273
                    /note="putative"
     polyA_site     3285
                    /note="putative"
ORIGIN

Alignment Scores:
Pred. No.:              8.21e-146      Length:        3235
Score:                  1818.50        Matches:       402
Percent Similarity:     59.57%         Conservative:  152
Best Local Similarity:  43.23%         Mismatches:    267
Query Match:            39.46%         Indels:        109
DB:                     4              Gaps:          17

US-10-768-511-8 (1-901) x AK089999 (1-3285)

Qy        14 SerProTyrProSerGlnSerGluGlyValSerThrProLeuProGlnValThrSerPro 33
Db       110 AGCAACCACAGCACCATGTCGTCCCCGGCATCCACCCCG--------------------- 148

Qy        34 SerPheAspAsnAlaAlaSerProValAlaGlyArgArgAlaValArgGlnThrProThr 53
Db       149 -----------------AGCCGCCGCAGCAGCCGACGCGGA---CGAGTCACCCCAACC 187

Qy        54 SerAlaValArgArgArgGlyArgGluThrAspSerAlaArgArgArgArgSerArgSer 73
Db       188 CAGTCCCTTCGAAGTGAGGAAAGCAGGTCGTCACCCAATCGGAGACGTAGAGGCGAAGAT 247

Qy        74 ArgSerLeuGly------------------------------------------------ 77
Db       248 TCTTCCACCGGAGAGCTACTGCCAATGCCCACCTCACCAGGAGCCGACCTGCAGAGCCCA 307

Qy        78 ---------AsnSerValTyrSerSerPro----------------------------- 84
Db       308 CCTGCACAGAATGCCTTGTTTTCCAGCCCTCCTCAGATGCATTCTTTAGCTATTCCTTTG 367

Qy        85 ---TyrAspAlaGlyThrProGlyThrProGlyThrProValAlaThrProValTyrAla 103
Db       368 GACTTTGATGTTAGTTCACCATTGACA--------------------------TATGGC 400

Qy       104 ThrProValGlyThrProMetGlyThrPro---SerPheHisArgGlyThrProGlnTyr 122
Db       401 ACTCCCAGCTCGCGAGTGGAAGGAACCCCAAGAAGTGGGGTGAGAGGCACACCT---GTA 457

Qy       123 LysGlnArgSerGluLeuGlySerGlnGlyLysProLeuHisArgArgArgArgSerGln 142
Db       458 AGGCAGAGGCCAGATCTGGGCTCAGCACGAAAGGGTTTGCAGGTGGATCTGCAGTCT--- 514

Qy       143 SerArgGluProGlyHisArgSerProSerArgGluProSerAlaAspGlyArgProSer 162
Db       515 ------------------------------------------GATGGCGCAGCAGCA 529

Qy       163 GluSerAlaGluProAspAspThrLeuGlyGlyGluTyrAlaTyrValTrpGlyThrAsn 182
Db       530 GAAGACATCGTACCAAGTGAACAGTCTCTAGGCCAAAAGCTTGTGATTTGGGGAACAGAT 589

Qy       183 ValAsnIleProAspValLeuArgAlaIleArgArgPheLeuHisAsnTyr--------- 199
Db       590 GTGAATGTGGCAACATGTAAAGAGAATTTTCAGAGATTCCTTCAGTGTTTTACTGATCCT 649

Qy       200 ---------------ArgSerSerAlaHisAspLeuAsnSerLys---TyrIleGlnIle 213
Db       650 CTGGCCCAAGAAGAAGAAAATGTTGGCATAGATATTACTCAACCTTTGTACATGCAACAA 709

Qy       214 IleGluGluThrValGluArgGluGluAspThrLeuAsnIleAspMetSerAspIleTyr 233
Db       710 CTTGGAGAGATTAATATTACAGGAGAGCCATTTTTAAATGTGAACTGCGAACACATAAAA 769

Qy       234 AspHisAspProAspLeuTyrAlaLysIleValArgTyrProLeuAspIleIleProLeu 253
Db       770 TCATTTAGCAAAAATCTGTATAGACAGCTCATCTCCTACCCACAGGAGGTTATACCAACC 829

Qy       254 LeuAspThrGluCysGlnGluValAlaThrSerLeuLeuPro------ThrPheGluLys 271
Db       830 TTTGACATGGCTGTCAATGAGATCTTCTTTGACCGTTATCCTGACTCCATCTTAGAACAT 889

Qy       272 HisIleGluAlaArgProPheAsnLeuLysAlaSerValHisMetArgGluLeuAsnPro 291
Db       890 CAGATTCAAGTCAGACCTTTTAATGCGTTGAAGACAAAGAGTATGAGAAACTTGAATCCA 949

Qy       292 SerAspIleAspLysLeuValSerValLysGlyMetValIleArgCysSerSerIleIle 311
Db       950 GAAGACATTGATCAGCTCATCACCATCAGTGGCATGGTCATCAGAACATCACAGCTGATT 1009

Qy       312 ProGluIleLysGlyAlaPhePheLysCysLeuValCysGlyHisSerProProLeuVal 331
Db      1010 CCGGAGATGCAGGAGGCCTTTTTCCAATGCCAAGTCTGTGCCCACACC---ACCCGGGTG 1066

Qy       332 ThrValValLysGlyArgValGluGluProThrArgCysGluLysProGluCysAlaAla 351
Db      1067 GAGATAGATCGAGGCAGAATTGCTGAGCCCTGCAGTTGT------GTGCACTGCCACACT 1120

Qy       352 ArgAsnAlaMetSerLeuIleHisAsnArgCysThrPheAlaAsnLysGlnIleValArg 371
Db      1121 ACCCACAGCATGGCACTGATCCACAACCGATCATTCTTCTCTGACAAGCAAATGATCAAA 1180

Qy       372 LeuGlnGluThrProAspAlaIleProGluGlyGluThrProHisThrValSerMetCys 391
Db      1181 CTTCAAGAGTCTCCTGAAGACATGCCTGCTGGGCAGACACCTCACACTATTGTCCTTTTT 1240

Qy       392 LeuTyrAsnThrMetValAspAlaValLysProGlyAspArgIleGluValThrGlyVal 411
Db      1241 GCCCACAATGACCTTGTTGACAAGGTTCAACCAGGGGACAGAGTGAACGTCACAGGCATA 1300

Qy       412 PheLysAlaMetAlaValArgValGlyProAsnGlnArgThrLeuArgAlaLeuTyrLys 431
Db      1301 TATCGAGCAGTACCAATTCGAGTTAATCCAAGAGTGAGCAACGTGAAGTCTGTCTATAAA 1360

Qy       432 ThrTyrIleAspCysValHisValLysLysSerAspArgGlyArgLeuGlnThrGluAsp 451
Db      1361 ACCCACATTGATGTCATTCATTATCGGAAAACGGATGCAAAACGTCTGCAT--------- 1411

Qy       452 ProMetGluMetAspLysGluAsnAspMetTyrAlaGlyTyrHisGluSerAspThrSer 471
Db      1412 ------------------------------------GGCCTTGATGAAGAAGCAGAACAG 1435

Qy       472 GluAlaAlaAsnGluAlaLysIleGlnLysLeuLysGluLeuSerLysLeuProGlyIle 491
Db      1436 AAACTTTTTTCAGAGAAACGTGTGAAATTGCTTAAGGAACTTTCCAGGAAGCCAGATATT 1495

Qy       492 TyrAspArgLeuSerArgSerLeuAlaProSerIleTrpGluLeuGluAspIleLysLys 511
Db      1496 TATGAGCGGCTTGCTTCAGCCTTGGCTCCCAGCATTTATGAACATGAAGATATCAAAAAG 1555

Qy       512 GlyLeuLeuCysGlnLeuPheGlyGlyLysAlaLysLysIle------ProSerGlyAla 529
Db      1556 GGAATCTTACTTCAGCTCTTTGGTGGAACAAGGAAGGATTTCAGTCACACTGGGAGGGGT 1615

Qy       530 SerPheArgGlyAspIleAsnValLeuLeuValGlyAspProGlyThrSerLysSerGln 549
Db      1616 AAATTCCGTGCTGAGATCAACATCCTTCTGTGTGGGGACCCTGGCACCAGCAAGTCCCAG 1675

Qy       550 LeuLeuGlnTyrValHisLysIleAlaProArgGlyIleTyrThrSerGlyArgGlySer 569
Db      1676 CTGCTACAGTATGTGTACAACCTGGTCCCCAGAGGCCAGTACACGTCTGGAAAAGGCTCC 1735
```

```
Qy        570 SerAlaValGlyLeuThrAlaTyrValThrLysAspProGluThrArgGluThrValLeu 589
              ||||||||||||||||||||||||||||     |||||||||||||||||:::   ||||||
Db       1736 AGTGCGGTCGGCCTCACCGCCTATGTGATGAAAGACCCTGAGACCAGGCAGCTTGTCCTC 1795

Qy        590 GluSerGlyAlaLeuValLeuSerAspArgGlyIleCysCysIleAspGluPheAspLys 609
              ::::::|||||||||||||||||||||   |||||||||||||||||:::||||||||||
Db       1796 CAGACAGGTGCCCTCGTCCTGAGTGACAATGGGATATGCTGCATCGATAAGTTTGACAAA 1855

Qy        610 MetSerAspAsnAlaArgSerMetLeuHisGluValMetGluGlnGlnThrValSerVal 629
              |||:::::::::   ||||||:::|||||||||||||||||||||||||||:::|||:::
Db       1856 ATGAATGAAAGCACAAGGTCTGTGCTGCATGAGGTCATGGAACAGCAGACTCTGTCCATT 1915

Qy        630 AlaLysGlyGlyIleIleAlaSerLeuAsnAlaArgThrSerValLeuAlaCysAlaAsn 649
              ||||||  ||||||||     ||||||||||||||||||||||||||||||    ||||||
Db       1916 GCAAAGGCTGGGATCATCTGTCAGCTCAATGCGCGCACCTCTGTCCTGGCAGCAGCAAAT 1975

Qy        650 ProSerGlySerArgTyrAsnAlaArgLeuSerValIleAspAsnIleGlnLeuProPro 669
              |||      |||::::::|||   :::    :::   |||:::||||||||||||||||
Db       1976 CCTATTGAGTCTCAGTGGAATCCTAAAAAAACAACCATTGAAAATATCCAACTACCGCAC 2035

Qy        670 ThrLeuLeuSerArgPheAspLeuIleTyrLeuMetLeuAspLysProAspGluGlnAsn 689
              |||||||||||||||||||||||||||:::|||||||||||      ||||||
Db       2036 ACATTGTTGTCAAGGTTTGATCTCATTTTCCTCATGCTAGACCCTCAGGATGAGGCATAT 2095

Qy        690 AspArgArgLeuAlaArgHisLeuValAlaLeuHisTyrGluAsnTyrGluValSerLys 709
              |||||||||||||||   |||||||||:::|||:::|||::::::   |||      :::
Db       2096 GACCGGCGTCTAGCTCATCACCTGGTTTCATTGTACTACCAAAGTGAGGAGCAAGTGGAG 2155

Qy        710 GlnAspAlaLeuAspLeuGlnThrLeuThrAlaTyrIleThrTyrAlaArgGlnHisVal 729
              ::::::   ||||||:::      |||      ||||||   ||||||          :::
Db       2156 GAGGAGTTCCTGGACATGGCCGTGCTGAAAGACTACATTGCATATGCCCATAGTACCATC 2215

Qy        730 HisProThrLeuSerAspGluAlaAlaGluAspLeuIleAsnGlyTyrValGluMetArg 749
              |||   ||||||:::||||||:::::   ||||||      ||||||    ||||||
Db       2216 ATGCCCCGACTGAGTGAGGAGGCCAGCCAGGCTCTCATTGAGGCTTATGTAAACATGAGG 2275

Qy        750 GlnLysGlyAsnPheProGlySerSerLysLysValIleThrAlaThrProArgGlnLeu 769
              :::               |||||||||:::   :::::::::|||   |||||||||||
Db       2276 AAG------------ATTGGGAGTAGCCGGGGGATGGTTTCTGCTTACCCTCGACAGCTA 2323

Qy        770 GluSerMetIleArgIleSerGluAlaLeuAlaArgMetArgPheSerGluValValGlu 789
              ||||||:::||||||:::::::||||||   |||::::::||||||||        ||||||
Db       2324 GAGTCATTAATTCGCTTAGCAGAAGCCCATGCTAAAGTAAGATTTTCAAACAAAGTTGAA 2383

Qy        790 LysValAspAlaAlaGluAlaValArgLeuLeuAspValAlaLeuGlnGlnSerAlaThr 809
                 :::|||      ||||||   ||||||         ||||||:::||||||||||||
Db       2384 GCAATTGATGTGGAAGAGGCAAAACGCCTCCACCGGGAGGCTCTGAAGCAGTCTGCAACT 2443

Qy        810 AspHisAlaThrGlyThrIleAspMetAspLeuIleThrThrGlyValSerAlaSerGlu 829
              |||      ||||||   :::|||:::      ::::::|||||||||:::||||||:::
Db       2444 GACCCTCGTACTGGCATGGTGGATATTTCTATTCTTACTACAGGAATGAGTGCCACTTCT 2503

Qy        830 ArgIleArgArgAlaAsnLeuLeuAlaAlaLeuArgGluLeuIleAlaAspLysIleSer 849
              |||   |||:::      |||      |||||||||:::||||||          :::|||
Db       2504 CGTAAACGGAAAGAAGAATTAGCTGAAGCATTGAGAAAACTTATT---------TTATCT 2554

Qy        850 ProGlySerSerSerGlyLeuLysThrSerGlnLeuLeuGluAspIleArgSerGlnSer 869
                 |||   :::      ||||||      ||||||   ||||||||||||   ||||||
Db       2555 AAGGGTAAAACACCAGCCTTAAAGTACCAACAGCTGTTTGAGGATATTCGGGGACAGTCT 2614

Qy        870 SerValAspValSerLeuGlnAspIleLysAsnAlaLeuGlySerLeuGlnGlyGluGly 889
                       ::::::            :::   ||||||   :::|||         :::
Db       2615 GACACAGCAATTACCAAGGACATGTTTGAAGAAGCCCTGCGAGCTTTGGCTGATGATGAT 2674

Qy        890 PheLeuThrValHisGlyAspIleValLys 899
              ||||||||||||   |||      |||:::
Db       2675 TTCCTAACAGTGACTGGGAAGACTGTCCGC 2704
```

RESULT 5
AK036653

```
LOCUS       AK036653                3545 bp    mRNA    linear   HTC 03-APR-2004
DEFINITION  Mus musculus adult male bone cDNA, RIKEN full-length enriched
            library, clone:9830146H02 product:mini chromosome maintenance
            deficient 4 homolog (S. cerevisiae), full insert sequence.
ACCESSION   AK036653
VERSION     AK036653.1  GI:26085352
KEYWORDS    HTC; CAP trapper.
SOURCE      Mus musculus (house mouse)
  ORGANISM  Mus musculus
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Euarchontoglires; Glires; Rodentia;
            Sciurognathi; Muridae; Murinae; Mus.
REFERENCE   1
  AUTHORS   Carninci,P. and Hayashizaki,Y.
  TITLE     High-efficiency full-length cDNA cloning
  JOURNAL   Meth. Enzymol. 303, 19-44 (1999)
   PUBMED   10349636
REFERENCE   2
  AUTHORS   Carninci,P., Shibata,Y., Hayatsu,N., Sugahara,Y., Shibata,K.,
            Itoh,M., Konno,H., Okazaki,Y., Muramatsu,M. and Hayashizaki,Y.
  TITLE     Normalization and subtraction of cap-trapper-selected cDNAs to
            prepare full-length cDNA libraries for rapid discovery of new genes
  JOURNAL   Genome Res. 10 (10), 1617-1630 (2000)
   PUBMED   11042159
REFERENCE   3
  AUTHORS   Shibata,K., Itoh,M., Aizawa,K., Nagaoka,S., Sasaki,N., Carninci,P.,
            Konno,H., Akiyama,J., Nishi,K., Kitsunai,T., Tashiro,H., Itoh,M.,
            Sumi,N., Ishii,Y., Nakamura,S., Hazama,M., Nishine,T., Harada,A.,
            Yamamoto,R., Matsumoto,H., Sakaguchi,S., Ikegami,T., Kashiwagi,K.,
            Fujiwake,S., Inoue,K., Togawa,Y., Izawa,M., Ohara,E., Watahiki,M.,
            Yoneda,Y., Ishikawa,T., Ozawa,K., Tanaka,T., Matsuura,S., Kawai,J.,
            Okazaki,Y., Muramatsu,M., Inoue,Y., Kira,A. and Hayashizaki,Y.
  TITLE     RIKEN integrated sequence analysis (RISA) system--384-format
            sequencing pipeline with 384 multicapillary sequencer
  JOURNAL   Genome Res. 10 (11), 1757-1771 (2000)
   PUBMED   11076861
REFERENCE   4
  AUTHORS   The RIKEN Genome Exploration Research Group Phase II Team and the
            FANTOM Consortium.
  TITLE     Functional annotation of a full-length mouse cDNA collection
  JOURNAL   Nature 409, 685-690 (2001)
REFERENCE   5
  AUTHORS   The FANTOM Consortium and the RIKEN Genome Exploration Research
            Group Phase I & II Team.
  TITLE     Analysis of the mouse transcriptome based on functional annotation
            of 60,770 full-length cDNAs
  JOURNAL   Nature 420, 563-573 (2002)
REFERENCE   6  (bases 1 to 3545)
  AUTHORS   Adachi,J., Aizawa,K., Akimura,T., Arakawa,T., Bono,H., Carninci,P.,
            Fukuda,S., Furuno,M., Hanagaki,T., Hara,A., Hashizume,W.,
            Hayashida,K., Hayatsu,N., Hiramoto,K., Hiraoka,T., Hirozane,T.,
            Hori,F., Imotani,K., Ishii,Y., Itoh,M., Kagawa,I., Kasukawa,T.,
            Katoh,H., Kawai,J., Kojima,Y., Kondo,S., Konno,H., Kouda,M.,
            Koya,S., Kurihara,C., Matsuyama,T., Miyazaki,A., Murata,M.,
            Nakamura,M., Nishi,K., Nomura,K., Numazaki,R., Ohno,M., Ohsato,N.,
            Okazaki,Y., Saito,R., Saitoh,H., Sakai,C., Sakai,K., Sakazume,N.,
            Sano,H., Sasaki,D., Shibata,K., Shinagawa,A., Shiraki,T.,
            Sogabe,Y., Tagami,M., Tagawa,A., Takahashi,F., Takaku-Akahira,S.,
            Takeda,Y., Tanaka,T., Tomaru,A., Toya,T., Yasunishi,A.,
            Muramatsu,M. and Hayashizaki,Y.
  TITLE     Direct Submission
  JOURNAL   Submitted (16-JUL-2001) Yoshihide Hayashizaki, The Institute of
            Physical and Chemical Research (RIKEN), Laboratory for Genome
            Exploration Research Group, RIKEN Genomic Sciences Center (GSC),
            RIKEN Yokohama Institute; 1-7-22 Suehiro-cho, Tsurumi-ku, Yokohama,
            Kanagawa 230-0045, Japan (E-mail:genome-res@gsc.riken.jp,
            URL:http://genome.gsc.riken.jp/, Tel:81-45-503-9222,
            Fax:81-45-503-9216)
COMMENT     cDNA library was prepared and sequenced in Mouse Genome
            Encyclopedia Project of Genome Exploration Research Group in Riken
            Genomic Sciences Center and Genome Science Laboratory in RIKEN.
            Division of Experimental Animal Research in Riken contributed to
            prepare mouse tissues.
            Please visit our web site for further details.
            URL:http://genome.gsc.riken.jp/
```

```
            URL:http://fantom.gsc.riken.jp/.
FEATURES             Location/Qualifiers
     source          1. .3545
                     /organism="Mus musculus"
                     /mol_type="mRNA"
                     /strain="C57BL/6J"
                     /db_xref="FANTOM_DB:9830146H02"
                     /db_xref="taxon:10090"
                     /clone="9830146H02"
                     /sex="male"
                     /tissue_type="bone"
                     /clone_lib="RIKEN full-length enriched mouse cDNA library"
                     /dev_stage="adult"
     misc_feature    180. .2769
                     /note="mini chromosome maintenance deficient 4 homolog (S.
                     cerevisiae) (MGD|MGI:103199, GB|NM_008565, evidence:
                     BLASTN, 100%, match=3221)
                     putative"
     polyA_signal    3525. .3530
                     /note="putative"
     polyA_site      3545
                     /note="putative"
ORIGIN

Alignment Scores:
Pred. No.:              5.51e-145     Length:        3545
Score:                  1809.50       Matches:       403
Percent Similarity:     59.51%        Conservative:  151
Best Local Similarity:  43.29%        Mismatches:    267
Query Match:            39.26%        Indels:        110
DB:                     4             Gaps:          17

US-10-768-511-8 (1-901) x AK036653 (1-3545)

Qy         14 SerProTyrProSerGlnSerGluGlyValSerThrProLeuProGlnValThrSerPro 33
              |||    :::   :::   |||            |||||||||
Db        165 AGCAACCACAGCACCATGTCGTCCCCGGCATCCACCCCG--------------------- 203

Qy         34 SerPheAspAsnAlaAlaSerProValAlaGlyArgArgAlaValArgGlnThrProThr 53
                                |||      :::   ||||||     |||    |||||||||
Db        204 ------------------AGCCGCCGCAGCAGCCGACGCGGA---CGAGTCACCCCAACC 242

Qy         54 SerAlaValArgArgArgGlyArgGluThrAspSerAlaArgArgArgArgSerArgSer 73
                 ::::::|||                   :::         |||||||||||
Db        243 CAGTCCCTTCGAAGTGAGGAAAGCAGGTCGTCACCCAATCGGAGACGTAGAGGCGAAGAT 302

Qy         74 ArgSerLeuGly------------------------------------------------ 77
                 |||   |||
Db        303 TCTTCCACCGGAGAGCTACTGCCAATGCCCACCTCACCAGGAGCCGACCTGCAGAGCCCA 362

Qy         78 ---------AsnSerValTyrSerSerPro------------------------------ 84
                       |||:::::::::|||||||||
Db        363 CCTGCACAGAATGCCTTGTTTTCCAGCCCTCCTCAGATGCATTCTTTAGCTATTCCTTTG 422

Qy         85 ---TyrAspAlaGlyThrProGlyThrProGlyThrProValAlaThrProValTyrAla 103
                 :::|||      :::|||   |||                           |||
Db        423 GACTTTGATGTTAGTTCACCATTGACA---------------------------TATGGC 455

Qy        104 ThrProValGlyThrProMetGlyThrPro---SerPheHisArgGlyThrProGlnTyr 122
              ||||||                  |||||||||   |||      | ||||||||||
Db        456 ACTCCCAGCTCGCGAGTGGAAGGAACCCCAAGAAGTGGGGTGAGAGGCACACCT---GTA 512

Qy        123 LysGlnArgSerGluLeuGlySer-GlnGlyLysProLeuHisArgArgArgArgSerGl 142
              :::||||||   :::|||||||||      |||   |||                :::|||
Db        513 AGGCAGAGGCCAGATCTGGGCTCAGCCACGAAAGGGTTTGCAGGTGGATCTGCAGTCT-- 570

Qy        142 nSerArgGluProGlyHisArgSerProSerArgGluProSerAlaAspGlyArgProSe 162
                                                               ||||||    ::
Db        571 ---------------------------------------------GATGGCGCAGCAGC 584

Qy        162 rGluSerAlaGluProAspAspThrLeuGlyGlyGluTyrAlaTyrValTrpGlyThrAs 182
               :|||         |||   :::         |||:::         :::|||||||||::
Db        585 AGAAGACATCGTACCAAGTGAACAGTCTCTAGGCCAAAAGCTTCTGATTTGGGGAACAGA 644

Qy        182 nValAsnIleProAspValLeuArgAlaIleArgArgPheLeuHisAsnTyr-------- 199
              :||||||:::                      :::||||||||||         :::
Db        645 TGTGAATGTGGCAACATGTAAAGAGAATTTTCAGAGATTCCTTCAGTGTTTTACTGATCC 704

Qy        200 ---------------ArgSerSerAlaHisAspLeuAsnSerLys---TyrIleGlnIl 213
                                    :::       |||:::              |||:::|||
Db        705 TCTGGCCAAAGAAGAAGAAAATGTTGGCATAGATATTACTCAACCTTTGTACATGCAACA 764

Qy        213 eIleGluGluThrValGluArgGluGluAspThrLeuAsnIleAspMetSerAspIleTy 233
               :::   |||                  |||      ||||||::::::          |||
Db        765 ACTTGGAGAGATTAATATTACAGGAGAGCCATTTTTAAATGTGAACTGCGAACACATAAA 824

Qy        233 rAspHisAspProAspLeuTyrAlaLysIleValArgTyrProLeuAspIleIleProLe 253
                          :::||||||   :::::::::      ||||||    ::::::||||||
Db        825 ATCATTTAGCAAAAATCTGTATAGACAGCTCATCTCCTACCCACAGGAGGTTATACCAAC 884

Qy        253 uLeuAspThrGluCysGlnGluValAlaThrSerLeuLeuPro------ThrPheGluLy 271
                  |||          |||:::                    |||              |||
Db        885 CTTTGACATGGCTGTCAATGAGATCTTCTTTGACCGTTATCCTGACTCCATCTTAGAACA 944

Qy        271 sHisIleGluAlaArgProPheAsnLeuLysAlaSerValHisMetArgGluLeuAsnPr 291
                  |||:::   ||||||||||||          :::      ||||||   |||||||||
Db        945 TCAGATTCAAGTCAGACCTTTTAATGCGTTGAAGACAAAGAGTATGAGAAACTTGAATCC 1004

Qy        291 oSerAspIleAspLysLeuValSerValLysGlyMetValIleArgCysSerSerIleIl 311
              |   ||||||||:::|||:::::::::   |||||||||||||||   |||     :::||
Db       1005 AGAAGACATTGATCAGCTCATCACCATCAGTGGCATGGTCATCAGAACATCACAGCTGAT 1064

Qy        311 eProGluIleLysGlyAlaPhePheLysCysLeuValCysGlyHisSerProProLeuVa 331
              ||||||:::::   |||||||||:::|||      ||||||   |||:::            ||
Db       1065 TCCGGAGATGCAGGAGGCCTTTTTCCAATGCCAAGTCTGTGCCCACACC---ACCCGGGT 1121

Qy        331 lThrValValLysGlyArgValGluGluProThrArgCysGluLysProGluCysAlaAl 351
              |   :::   :::||||||:::      ||||||         |||            |||
Db       1122 GGAGATAGATCGAGGCAGAATTGCTGAGCCCTGCAGTTGT------GTGCACTGCCACAC 1175

Qy        351 aArgAsnAlaMetSerLeuIleHisAsnArgCysThrPheAlaAsnLysGlnIleValAr 371
                ::::::|||:::|||||||||||||||       |||::::::||||||:::::::::
Db       1176 TACCCACAGCATGGCACTGATCCACAACCGATCATTCTTCTCTGACAAGCAAATGATCAA 1235

Qy        371 gLeuGlnGluThrProAspAlaIleProGluGlyGluThrProHisThrValSerMetCy 391
               :|||||||||:::|||:::     :::|||   |||:::|||||||||||||:::    :::
Db       1236 ACTTCAAGAGTCTCCTGAAGACATGCCTGCTGGGCAGACACCTCACACTATTGTCCTTTT 1295

Qy        391 sLeuTyrAsnThrMetValAspAlaValLysProGlyAspArgIleGluValThrGlyVa 411
                 :::|||   :::||||||   |||:::||||||||||||:::    |||||||||::
Db       1296 TGCCCACAATGACCTTGTTGACAAGGTTCAACCAGGGGACAGAGTGAACGTCACAGGCAT 1355

Qy        411 lPheLysAlaMetAlaValArgValGlyProAsnGlnArgThrLeuArgAlaLeuTyrLy 431
              ::::::|||:::   :::||||||   |||               :::::::::::||||
Db       1356 ATATCGAGCAGTACCAATTCGAGTTAATCCAAGAGTGAGCAACGTGAAGTCTGTCTATAA 1415

Qy        431 sThrTyrIleAspCysValHisValLysLysSerAspArgGlyArgLeuGlnThrGluAs 451
              |||:::||||||   :::|||   :::|||:::|||      ||||||
Db       1416 AACCCACATTGATGTCATTCATTATCGGAAAACGGATGCAAAACGTCTGCAT-------- 1467

Qy        451 pProMetGluMetAspLysGluAsnAspMetTyrAlaGlyTyrHisGluSerAspThrSe 471
                                                   |||      |||
Db       1468 -------------------------------------GGCCTTGATGAAGAAGCAGAACA 1490

Qy        471 rGluAlaAlaAsnGluAlaLysIleGlnLysLeuLysGluLeuSerLysLeuProGlyIl 491
               :::      :::|||   :::::::::   ||||||||||||||||:::   |||    ||
Db       1491 GAAACTTTTTTCAGAGAAACGTGTGAAATTGCTTAAGGAACTTTCCAGGAAGCCAGATAT 1550

Qy        491 eTyrAspArgLeuSerArgSerLeuAlaProSerIleTrpGluLeuGluAspIleLysLy 511
              ||||:::||||||:::   :::||||||||||||||||:::|||   ||||||||||||
Db       1551 TTATGAGCGGCTTGCTTCAGCCTTGGCTCCCAGCATTTATGAACATGAAGATATCAAAAA 1610

Qy        511 sGlyLeuLeuCysGlnLeuPheGlyGlyLysAlaLysLysIle------ProSerGlyAl 529
              ||||:::|||   ||||||||||||||               |||
Db       1611 GGGAATCTTACTTCAGCTCTTTGGTGGAACAAGGAAGGATTTCAGTCACACTGGGAGGGG 1670
```

```
Qy        529 aSerPheArgGlyAspIleAsnValLeuLeuValGlyAspProGlyThrSerLysSerGl 549
                 |||||    :::||||||:::||||||   ||||||||||||||||||||||||
Db       1671 TAAATTCCGTGCTGAGATCAACATCCTTCTGTGTGGGGACCCTGGCACCAGCAAGTCCCA 1730

Qy        549 nLeuLeuGlnTyrValHisLysIleAlaProArgGlyIleTyrThrSerGlyArgGlySe 569
              |||||||||||||||||:::   :::   ||||||||   ||||||||||||:::|||||
Db       1731 GCTGCTACAGTATGTGTACAACCTGGTCCCCAGAGGCCAGTACACGTCTGGAAAAGGCTC 1790

Qy        569 rSerAlaValGlyLeuThrAlaTyrValThrLysAspProGluThrArgGluThrValLe 589
              ||||||||||||||||||||||||||||||   ||||||||||||||||:::   |||||
Db       1791 CAGTGCGGTCGGCCTCACCGCCTATGTGATGAAAGACCCTGAGACCAGGCAGCTTGTCCT 1850

Qy        589 uGluSerGlyAlaLeuValLeuSerAspArgGlyIleCysCysIleAspGluPheAspLy 609
              |::::::|||||||||||||||||||||   |||||||||||||||||||||||||||||
Db       1851 CCAGACAGGTGCCCTCGTCCTGAGTGACAATGGGATATGCTGCATCGATGAGTTTGACAA 1910

Qy        609 sMetSerAspAsnAlaArgSerMetLeuHisGluValMetGluGlnGlnThrValSerVa 629
              ||||:::::::::   ||||||:::|||||||||||||||||||||||||||:::|||::
Db       1911 AATGAATGAAAGCACAAGGTCTGTGCTGCATGAGGTCATGGAACAGCAGACTCTGTCCAT 1970

Qy        629 lAlaLysGlyGlyIleIleAlaSerLeuAsnAlaArgThrSerValLeuAlaCysAlaAs 649
              :||||||   |||||||||      |||||||||||||||||||||||||   |||||
Db       1971 TGCAAAGGCTGGGATCATCTGTCAGCTCAATGCGCGCACCTCTGTCCTGGCAGCAGCAAA 2030

Qy        649 nProSerGlySerArgTyrAsnAlaArgLeuSerValIleAspAsnIleGlnLeuProPr 669
              ||||   |||::::::|||     :::   :::   |||:::||||||||||||||||
Db       2031 TCCTATTGAGTCTCAGTGGAATCCTAAAAAAACAACCATTGAAAATATCCAACTACCGCA 2090

Qy        669 oThrLeuLeuSerArgPheAspLeuIleTyrLeuMetLeuAspLysProAspGluGlnAs 689
              ||||||||||||||||||||||||||||||:::||||||||||||   ||||||
Db       2091 CACATTGTTGTCAAGGTTTGATCTCATTTTCCTCATGCTAGACCCTCAGGATGAGGCATA 2150

Qy        689 nAspArgArgLeuAlaArgHisLeuValAlaLeuHisTyrGluAsnTyrGluValSerLy 709
              |||||||||||||||   ||||||||||:::|||:::|||::::::   |||      ::
Db       2151 TGACCGGCGTCTAGCTCATCACCTGGTTTCATTGTACTACCAAAGTGAGGAGCAAGTGGA 2210

Qy        709 sGlnAspAlaLeuAspLeuGlnThrLeuThrAlaTyrIleThrTyrAlaArgGlnHisVa 729
              :::::::   ||||||:::      |||      ||||||   ||||||         ::
Db       2211 GGAGGAGTTCCTGGACATGGCCGTGCTGAAAGACTACATTGCATATGCCCATAGTACCAT 2270

Qy        729 lHisProThrLeuSerAspGluAlaAlaGluAspLeuIleAsnGlyTyrValGluMetAr 749
              :   |||   ||||||:::||||||::::::   ||||||      ||||||   |||||
Db       2271 CATGCCCCGACTGAGTGAGGAGGCCAGCCAGGCTCTCATTGAGGCTTATGTAAACATGAG 2330

Qy        749 gGlnLysGlyAsnPheProGlySerSerLysLysValIleThrAlaThrProArgGlnLe 769
              |:::                |||||||:::   :::::::::|||   ||||||||||
Db       2331 GAAG------------ATTGGGAGTAGCCGGGGGATGGTTTCTGCTTACCCTCGACAGCT 2378

Qy        769 uGluSerMetIleArgIleSerGluAlaLeuAlaArgMetArgPheSerGluValValGl 789
              ||||||:::||||||:::::||||||   |||:::::::|||||||||      |||||
Db       2379 AGAGTCATTAATTCGCTTAGCAGAAGCCCATGCTAAAGTAAGATTTTCAAACAAAGTTGA 2438

Qy        789 uLysValAspAlaAlaGluAlaValArgLeuLeuAspValAlaLeuGlnGlnSerAlaTh 809
              |   :::|||      ||||||   ||||||         ||||||:::||||||||||||
Db       2439 AGCAATTGATGTGGAAGAGGCAAAACGCCTCCACCGGGAGGCTCTGAAGCAGTCTGCAAC 2498

Qy        809 rAspHisAlaThrGlyThrIleAspMetAspLeuIleThrThrGlyValSerAlaSerGl 829
              ||||      ||||||   :::|||:::   ::::::|||||||||:::|||||||:::
Db       2499 TGACCCTCGTACTGGCATTGTGGATATTTCTATTCTTACTACAGGAATGAGTGCCACTTC 2558

Qy        829 uArgIleArgArgAlaAsnLeuLeuAlaAlaLeuArgGluLeuIleAlaAspLysIleSe 849
                |||   |||:::      |||      |||||||||:::||||||         :::||
Db       2559 TCGTAAACGGAAAGAAGAATTAGCTGAAGCATTGAGAAAACTTATT---------TTATC 2609

Qy        849 rProGlySerSerSerGlyLeuLysThrSerGlnLeuLeuGluAspIleArgSerGlnSe 869
              |   |||   :::         ||||||      |||||||||||||         |||||
Db       2610 TAAGGGTAAAACACCAGCCTTAAAGTACCAACAGCTGTTTGAGGATATTCGGGGACAGTC 2669

Qy        869 rSerValAspValSerLeuGlnAspIleLysAsnAlaLeuGlySerLeuGlnGlyGluGl 889
              |         ::::::         :::      ||||||   :::|||      :::
Db       2670 TGACACAGCAATTACCAAGGACATGTTTGAAGAAGCCCTGCGAGCTTTGGCTGATGATGA 2729

Qy        889 yPheLeuThrValHisGlyAspIleValLys 899
```

```
                 |||||||||||   |||      |||:::
Db       2730 TTTCCTAACAGTGACTGGGAAGACTGTCCGC 2760

RESULT 6
AK011743
LOCUS       AK011743                3316 bp    mRNA    linear   HTC 03-APR-2004
DEFINITION  Mus musculus 10 days embryo whole body cDNA, RIKEN full-length
            enriched library, clone:2610042I15 product:mini chromosome
            maintenance deficient 4 homolog (S. cerevisiae), full insert
            sequence.
ACCESSION   AK011743
VERSION     AK011743.1  GI:12848060
KEYWORDS    HTC; CAP trapper.
SOURCE      Mus musculus (house mouse)
  ORGANISM  Mus musculus
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Euarchontoglires; Glires; Rodentia;
            Sciurognathi; Muridae; Murinae; Mus.
REFERENCE   1
  AUTHORS   Carninci,P. and Hayashizaki,Y.
  TITLE     High-efficiency full-length cDNA cloning
  JOURNAL   Meth. Enzymol. 303, 19-44 (1999)
   PUBMED   10349636
REFERENCE   2
  AUTHORS   Carninci,P., Shibata,Y., Hayatsu,N., Sugahara,Y., Shibata,K.,
            Itoh,M., Konno,H., Okazaki,Y., Muramatsu,M. and Hayashizaki,Y.
  TITLE     Normalization and subtraction of cap-trapper-selected cDNAs to
            prepare full-length cDNA libraries for rapid discovery of new genes
  JOURNAL   Genome Res. 10 (10), 1617-1630 (2000)
   PUBMED   11042159
REFERENCE   3
  AUTHORS   Shibata,K., Itoh,M., Aizawa,K., Nagaoka,S., Sasaki,N., Carninci,P.,
            Konno,H., Akiyama,J., Nishi,K., Kitsunai,T., Tashiro,H., Itoh,M.,
            Sumi,N., Ishii,Y., Nakamura,S., Hazama,M., Nishine,T., Harada,A.,
            Yamamoto,R., Matsumoto,H., Sakaguchi,S., Ikegami,T., Kashiwagi,K.,
            Fujiwake,S., Inoue,K., Togawa,Y., Izawa,M., Ohara,E., Watahiki,M.,
            Yoneda,Y., Ishikawa,T., Ozawa,K., Tanaka,T., Matsuura,S., Kawai,J.,
            Okazaki,Y., Muramatsu,M., Inoue,Y., Kira,A. and Hayashizaki,Y.
  TITLE     RIKEN integrated sequence analysis (RISA) system--384-format
            sequencing pipeline with 384 multicapillary sequencer
  JOURNAL   Genome Res. 10 (11), 1757-1771 (2000)
   PUBMED   11076861
REFERENCE   4
  AUTHORS   The RIKEN Genome Exploration Research Group Phase II Team and the
            FANTOM Consortium.
  TITLE     Functional annotation of a full-length mouse cDNA collection
  JOURNAL   Nature 409, 685-690 (2001)
REFERENCE   5
  AUTHORS   The FANTOM Consortium and the RIKEN Genome Exploration Research
            Group Phase I & II Team.
  TITLE     Analysis of the mouse transcriptome based on functional annotation
            of 60,770 full-length cDNAs
  JOURNAL   Nature 420, 563-573 (2002)
REFERENCE   6  (bases 1 to 3316)
  AUTHORS   Adachi,J., Aizawa,K., Akahira,S., Akimura,T., Arai,A., Aono,H.,
            Arakawa,T., Bono,H., Carninci,P., Fukuda,S., Fukunishi,Y.,
            Furuno,M., Hanagaki,T., Hara,A., Hayatsu,N., Hiramoto,K.,
            Hiraoka,T., Hori,F., Imotani,K., Ishii,Y., Itoh,M., Izawa,M.,
            Kasukawa,T., Kato,H., Kawai,J., Kojima,Y., Konno,H., Kouda,M.,
            Koya,S., Kurihara,C., Matsuyama,T., Miyazaki,A., Nishi,K.,
            Nomura,K., Numazaki,R., Ohno,M., Okazaki,Y., Okido,T., Owa,C.,
            Saito,H., Saito,R., Sakai,C., Sakai,K., Sano,H., Sasaki,D.,
            Shibata,K., Shibata,Y., Shinagawa,A., Shiraki,T., Sogabe,Y.,
            Suzuki,H., Tagami,M., Tagawa,A., Takahashi,F., Tanaka,T.,
            Tejima,Y., Toya,T., Yamamura,T., Yasunishi,A., Yoshida,K.,
            Yoshino,M., Muramatsu,M. and Hayashizaki,Y.
  TITLE     Direct Submission
  JOURNAL   Submitted (10-JUL-2000) Yoshihide Hayashizaki, The Institute of
            Physical and Chemical Research (RIKEN), Laboratory for Genome
            Exploration Research Group, RIKEN Genomic Sciences Center (GSC),
            RIKEN Yokohama Institute; 1-7-22 Suehiro-cho, Tsurumi-ku, Yokohama,
            Kanagawa 230-0045, Japan (E-mail:genome-res@gsc.riken.jp,
            URL:http://genome.gsc.riken.jp/, Tel:81-45-503-9222,
```

```
            Fax:81-45-503-9216)
COMMENT     Please visit our web site (http://genome.gsc.riken.jp/) for further
            details.
            cDNA library was prepared and sequenced in Mouse Genome
            Encyclopedia Project of Genome Exploration Research Group in Riken
            Genomic Sciences Center and Genome Science Laboratory in RIKEN.
            Division of Experimental Animal Research in Riken contributed to
            prepare mouse tissues. First strand cDNA was primed with a primer
            [5' GAGAGAGAGAAGGATCCAAGAGCTCTTTTTTTTTTTTTTTTVN 3'], cDNA was
            prepared by using trehalose thermo-activated reverse transcriptase
            and subsequently enriched for full-length by cap-trapper. Second
            strand cDNA was prepared with the primer adapter of sequence [5'
            GAGAGAGAGATTCTCGAGTTAATTAAATTAATCCCCCCCCCCCC 3']. cDNA was cleaved
            with XhoI and SstI. Cloning sites, 5' end: XhoI; 3' end: SstI.
            Host: SOLR.
FEATURES             Location/Qualifiers
     source          1..3316
                     /organism="Mus musculus"
                     /mol_type="mRNA"
                     /strain="C57BL/6J"
                     /db_xref="FANTOM_DB:2610042I15"
                     /db_xref="taxon:10090"
                     /clone="2610042I15"
                     /tissue_type="whole body"
                     /clone_lib="RIKEN full-length enriched mouse cDNA library"
                     /dev_stage="10 days embryo"
     CDS             156..2744
                     /note="unnamed protein product; mini chromosome
                     maintenance deficient 4 homolog (S. cerevisiae)
                     (MGD|MGI:103199, GB|NM_008565, evidence: BLASTN, 100%,
                     match=3221)
                     putative"
                     /codon_start=1
                     /protein_id="BAB27813.1"
                     /db_xref="GI:12848061"
                     /translation="MSSPASTPSRRSSRRGRVTPTQSLRSEESRSSPNRRRRGEDSST
                     GELLPMPTSPGADLQSPPAQNALFSSPPQMHSLAIPLDFDVSSPLTYGTPSSRVEGTP
                     RSGVRGTPVRQRPDLGSARKGLQVDLQSDGAAAEDIVPSEQSLGQKLVIWGTDVNVAT
                     CKENFQRFLQCFTDPLAKEEENVGIDITQPLYMQQLGEINITGEPFLNVNCEHIKSFS
                     KNLYRQLISYPQEVIRTFDMAVNEIFFDRYPDSILEHQIQVRPFNALKTKSMRNLNPE
                     DIDQLITISGMVIRTSQLIPEMQEAFFQCQVCAHTTRVEIDRGRIAEPCSCVHCHTTH
                     SMALIHNRSFFSDKQMIKLQESPEDMPAGQTPHTIVLFAHNDLVDKVQPGDRVNVAGI
                     YRAVPIRVNPRVSNVKSVYKTHIDVIHYRKTDAKRLHGLDEEAEQKLFSEKRVKLLKE
                     LSRKPDIYERLASALAPSIYEHEDIKKGILLQLFGGTRKDFSHTGRGKFRAEINILLC
                     GDPGTSKSQLLQYVYNLVPRGQYTSGKGSSAVGLTAYVMKDPETRQLVLQTGALVLSD
                     NGICCIDEFDKMNESTRSVLHEVMEQQTLSIAKAGIICQLNARTSVLAAANPIESQWN
                     PKKTTIENIQLPHTLLSRFDLIFLMLDPQDEAYDRRLAHHLVSLYYQSEEQVEEEFLD
                     MAVLKDYIAYAHSTIMPRLSEEASQALIEAYVNMRKIGSSRGMVSAYPRQLESLIRLA
                     EAHAKVRFSNKVEAIDVEEAKRLHREALKQSATDPRTGIVDISILTTGMSATSRKRKE
                     ELAEALRKLILSKGKTPALKYQQLFEDIRGQSDTAITKDMFEEALRALADDDFLTVTG
                     KTVRLL"
     polyA_signal    3299..3304
                     /note="putative"
     polyA_site      3316
                     /note="putative"
ORIGIN

Alignment Scores:
Pred. No.:              6.1e-145        Length:         3316
Score:                  1808.50         Matches:        401
Percent Similarity:     59.35%          Conservative:   151
Best Local Similarity:  43.12%          Mismatches:     269
Query Match:            39.24%          Indels:         109
DB:                     4               Gaps:           17

US-10-768-511-8 (1-901) x AK011743 (1-3316)

Qy         14 SerProTyrProSerGlnSerGluGlyValSerThrProLeuProGlnValThrSerPro 33
                |||   :::   :::   |||         ||||||||
Db        141 AGCAACCACAGCACCATGTCGTCCCCGGCATCCACCCCG--------------------- 179

Qy         34 SerPheAspAsnAlaAlaSerProValAlaGlyArgArgAlaValArgGlnThrProThr 53
                             |||      :::   ||||||      |||    |||||||||
Db        180 ------------------AGCCGCCGCAGCAGCCGACGCGGA---CGAGTCACCCCAACC 218

Qy         54 SerAlaValArgArgArgGlyArgGluThrAspSerAlaArgArgArgArgSerArgSer 73
                ::::::|||            :::          ||||||||||||
Db        219 CAGTCCCTTCGAAGTGAGGAAAGCAGGTCGTCACCCAATCGGAGACGTAGAGGCGAAGAT 278

Qy         74 ArgSerLeuGly------------------------------------------------ 77
                 |||   |||
Db        279 TCTTCCACCGGAGAGCTACTGCCAATGCCCACCTCACCAGGAGCCGACCTGCAGAGCCCA 338

Qy         78 ---------AsnSerValTyrSerSerPro------------------------------ 84
                          |||::::::::||||||||
Db        339 CCTGCACAGAATGCCTTGTTTTCCAGCCCTCCTCAGATGCATTCTTTAGCTATTCCTTTG 398

Qy         85 ---TyrAspAlaGlyThrProGlyThrProGlyThrProValAlaThrProValTyrAla 103
                 :::|||      :::|||   |||                          |||
Db        399 GACTTTGATGTTAGTTCACCATTGACA-------------------------TATGGC 431

Qy        104 ThrProValGlyThrProMetGlyThrPro---SerPheHisArgGlyThrProGlnTyr 122
                ||||||          |||||||||    |||      ||||||||||||
Db        432 ACTCCCAGCTCGCGAGTGGAAGGAACCCCAAGAAGTGGGGTGAGAGGCACACCT---GTA 488

Qy        123 LysGlnArgSerGluLeuGlySerGlnGlyLysProLeuHisArgArgArgArgSerGln 142
                :::||||||   :::|||||||||         |||   |||          :::|||
Db        489 AGGCAGAGGCCAGATCTGGGCTCAGCACGAAAGGGTTTGCAGGTGGATCTGCAGTCT--- 545

Qy        143 SerArgGluProGlyHisArgSerProSerArgGluProSerAlaAspGlyArgProSer 162
                                                        ||||||      :::
Db        546 ---------------------------------------------GATGGCGCAGCAGCA 560

Qy        163 GluSerAlaGluProAspAspThrLeuGlyGlyGluTyrAlaTyrValTrpGlyThrAsn 182
                |||         |||   :::         |||:::         :::|||||||||:::
Db        561 GAAGACATCGTACCAAGTGAACAGTCTCTAGGCCAAAAGCTTGTGATTTGGGGAACAGAT 620

Qy        183 ValAsnIleProAspValLeuArgAlaIleArgArgPheLeuHisAsnTyr--------- 199
                ||||||:::                 :::||||||||||||      :::
Db        621 GTGAATGTGGCAACATGTAAAGAGAATTTTCAGAGATTCCTTCAGTGTTTTACTGATCCT 680

Qy        200 ---------------ArgSerSerAlaHisAspLeuAsnSerLys---TyrIleGlnIle 213
                              :::         |||:::         |||:::|||
Db        681 CTGGCCAAAGAAGAAGAAAATGTTGGCATAGATATTACTCAACCTTTGTACATGCAACAA 740

Qy        214 IleGluGluThrValGluArgGluGluAspThrLeuAsnIleAspMetSerAspIleTyr 233
                :::   |||                  |||      ||||||::::::          |||
Db        741 CTTGGAGAGATTAATATTACAGGAGAGCCATTTTTAAATGTGAACTGCGAACACATAAAA 800

Qy        234 AspHisAspProAspLeuTyrAlaLysIleValArgTyrProLeuAspIleIleProLeu 253
                         :::||||||   :::::::::      ||||||      ::::::|||
Db        801 TCATTTAGCAAAAATCTGTATAGACAGCTCATCTCCTACCCACAGGAGGTTATACGAACC 860

Qy        254 LeuAspThrGluCysGlnGluValAlaThrSerLeuLeuPro------ThrPheGluLys 271
                   |||         |||:::                  |||
Db        861 TTTGACATGGCTGTCAATGAGATCTTCTTTGACCGTTATCCTGACTCCATCTTAGAACAT 920

Qy        272 HisIleGluAlaArgProPheAsnLeuLysAlaSerValHisMetArgGluLeuAsnPro 291
                   |||:::   ||||||||||||      :::         ||||||   |||||||||
Db        921 CAGATTCAAGTCAGACCTTTTAATGCGTTGAAGACAAAGAGTATGAGAAACTTGAATCCA 980

Qy        292 SerAspIleAspLysLeuValSerValLysGlyMetValIleArgCysSerSerIleIle 311
                   |||||||||:::|||:::::::::      ||||||||||||||   |||   :::|||
Db        981 GAAGACATTGATCAGCTCATCACCATCAGTGGCATGGTCATCAGAACATCACAGCTGATT 1040

Qy        312 ProGluIleLysGlyAlaPhePheLysCysLeuValCysGlyHisSerProProLeuVal 331
                ||||||::::::   |||||||||:::|||   ||||||   |||:::          |||
Db       1041 CCGGAGATGCAGGAGGCCTTTTTCCAATGCCAAGTCTGTGCCCACACC---ACCCGGGTG 1097

Qy        332 ThrValValLysGlyArgValGluGluProThrArgCysGluLysProGluCysAlaAla 351
                :::   :::|||||||:::         ||||||         |||            |||
Db       1098 GAGATAGATCGAGGCAGAATTGCTGAGCCCTGCAGTTGT------GTGCACTGCCACACT 1151

Qy        352 ArgAsnAlaMetSerLeuIleHisAsnArgCysThrPheAlaAsnLysGlnIleValArg 371
                ::::::|||:::|||||||||||||||      |||::::::||||||:::::::::
Db       1152 ACCCACAGCATGGCACTGATCCACAACCGATCATTCTTCTCTGACAAGCAAATGATCAAA 1211
```

```
Qy        372 LeuGlnGluThrProAspAlaIleProGluGlyGluThrProHisThrValSerMetCys 391
              ||||||||:::|||:::    :::|||   |||:::||||||||||||:::     :::
Db       1212 CTTCAAGAGTCTCCTGAAGACATGCCTGCTGGGCAGACACCTCACACTATTGTCCTTTTT 1271

Qy        392 LeuTyrAsnThrMetValAspAlaValLysProGlyAspArgIleGluValThrGlyVal 411
                 :::|||   :::||||||   |||:::|||||||||||||:::   |||    |||:::
Db       1272 GCCCACAATGACCTTGTTGACAAGGTTCAACCAGGGGACAGAGTGAACGTCGCAGGCATA 1331

Qy        412 PheLysAlaMetAlaValArgValGlyProAsnGlnArgThrLeuArgAlaLeuTyrLys 431
              ::::::|||:::   :::|||||   |||               ::::::::::::||||||
Db       1332 TATCGAGCAGTACCAATTCGAGTTAATCCAAGAGTGAGCAACGTGAAGTCTGTCTATAAA 1391

Qy        432 ThrTyrIleAspCysValHisValLysLysSerAspArgGlyArgLeuGlnThrGluAsp 451
              |||:::|||||   :::|||   :::|||:::|||         ||||||
Db       1392 ACCCACATTGATGTCATTCATTATCGGAAAACGGATGCAAAACGTCTGCAT--------- 1442

Qy        452 ProMetGluMetAspLysGluAsnAspMetTyrAlaGlyTyrHisGluSerAspThrSer 471
                                                   |||   |||
Db       1443 ------------------------------------GGCCTTGATGAAGAAGCAGAACAG 1466

Qy        472 GluAlaAlaAsnGluAlaLysIleGlnLysLeuLysGluLeuSerLysLeuProGlyIle 491
              :::      :::|||   :::::::::   ||||||||||||||:::   |||   |||
Db       1467 AAACTTTTTCAGAGAAACGTGTGAAATTGCTTAAGGAACTTTCCAGGAAGCCAGATATT 1526

Qy        492 TyrAspArgLeuSerArgSerLeuAlaProSerIleTrpGluLeuGluAspIleLysLys 511
              |||:::||||||:::   :::||||||||||||||||:::|||   ||||||||||||||
Db       1527 TATGAGCGGCTTGCTTCAGCCTTGGCTCCCAGCATTTATGAACATGAAGATATCAAAAAG 1586

Qy        512 GlyLeuLeuCysGlnLeuPheGlyGlyLysAlaLysLysIle------ProSerGlyAla 529
              |||:::|||   ||||||||||||||||||         |||
Db       1587 GGAATCTTACTTCAGCTCTTTGGTGGAACAAGGAAGGATTTCAGTCACACTGGGAGGGGT 1646

Qy        530 SerPheArgGlyAspIleAsnValLeuLeuValGlyAspProGlyThrSerLysSerGln 549
                 ||||||   :::||||||:::|||||||   |||||||||||||||||||||||||||
Db       1647 AAATTCCGTGCTGAGATCAACATCCTTCTGTGTGGGGACCCTGGCACCAGCAAGTCCCAG 1706

Qy        550 LeuLeuGlnTyrValHisLysIleAlaProArgGlyIleTyrThrSerGlyArgGlySer 569
              ||||||||||||||||:::   :::   |||||||||   ||||||||||||:::||||||
Db       1707 CTGCTACAGTATGTGTACAACCTGGTCCCCAGAGGCCAGTACACGTCTGGAAAAGGCTCC 1766

Qy        570 SerAlaValGlyLeuThrAlaTyrValThrLysAspProGluThrArgGluThrValLeu 589
              |||||||||||||||||||||||||||   ||||||||||||||||||:::   ||||||
Db       1767 AGTGCGGTCGGCCTCACCGCCTATGTGATGAAAGACCCTGAGACCAGGCAGCTTGTCCTC 1826

Qy        590 GluSerGlyAlaLeuValLeuSerAspArgGlyIleCysCysIleAspGluPheAspLys 609
              ::::::||||||||||||||||||||||||   |||||||||||||||||||||||||||
Db       1827 CAGACAGGTGCCCTCGTCCTGAGTGACAATGGGATATGCTGCATCGATGAGTTTGACAAA 1886

Qy        610 MetSerAspAsnAlaArgSerMetLeuHisGluValMetGluGlnGlnThrValSerVal 629
              |||::::::::::   ||||||:::||||||||||||||||||||||||:::|||:::
Db       1887 ATGAATGAAAGCACAAGGTCTGTGCTGCATGAGGTCATGGAACAGCAGACTCTGTCCATT 1946

Qy        630 AlaLysGlyGlyIleIleAlaSerLeuAsnAlaArgThrSerValLeuAlaCysAlaAsn 649
              ||||||   |||||||||   |||||||||||||||||||||||||||   ||||||
Db       1947 GCAAAGGCTGGGATCATCTGTCAGCTCAATGCGCGCACCTCTGTCCTGGCAGCAGCAAAT 2006

Qy        650 ProSerGlySerArgTyrAsnAlaArgLeuSerValIleAspAsnIleGlnLeuProPro 669
              |||   |||:::::::|||   :::   :::   |||:::|||||||||||||||
Db       2007 CCTATTGAGTCTCAGTGGAATCCTAAAAAAACAACCATTGAAAATATCCAACTACCGCAC 2066

Qy        670 ThrLeuLeuSerArgPheAspLeuIleTyrLeuMetLeuAspLysProAspGluGlnAsn 689
              |||||||||||||||||||||||||||:::||||||||||||      ||||||
Db       2067 ACATTGTTGTCAAGGTTTGATCTCATTTTCCTCATGCTAGACCCTCAGGATGAGGCATAT 2126

Qy        690 AspArgArgLeuAlaArgHisLeuValAlaLeuHisTyrGluAsnTyrGluValSerLys 709
              ||||||||||||||   ||||||||:::|||:::|||::::::   |||      :::
Db       2127 GACCGGCGTCTAGCTCATCACCTGGTTTCATTGTACTACCAAAGTGAGGAGCAAGTGGAG 2186

Qy        710 GlnAspAlaLeuAspLeuGlnThrLeuThrAlaTyrIleThrTyrAlaArgGlnHisVal 729
              ::::::   ||||||:::      |||      ||||||   ||||||         :::
Db       2187 GAGGAGTTCCTGGACATGGCCGTGCTGAAAGACTACATTGCATATGCCCATAGTACCATC 2246

Qy        730 HisProThrLeuSerAspGluAlaAlaGluAspLeuIleAsnGlyTyrValGluMetArg 749
                 |||   ||||||:::||||||::::::   ||||||      ||||||   ||||||
Db       2247 ATGCCCCGACTGAGTGAGGAGGCCAGCCAGGCTCTCATTGAGGCTTATGTAAACATGAGG 2306

Qy        750 GlnLysGlyAsnPheProGlySerSerLysLysValIleThrAlaThrProArgGlnLeu 769
              :::                |||||||||:::   :::::::::|||   ||||||||||||
Db       2307 AAG------------ATTGGGAGTAGCCGGGGGATGGTTTCTGCTTACCCTCGACAGCTA 2354

Qy        770 GluSerMetIleArgIleSerGluAlaLeuAlaArgMetArgPheSerGluValValGlu 789
              ||||||:::||||||:::::::||||||      |||:::::::|||||||||      ||||||
Db       2355 GAGTCATTAATTCGCTTAGCAGAAGCCCATGCTAAAGTAAGATTTTCAAACAAAGTTGAA 2414

Qy        790 LysValAspAlaAlaGluAlaValArgLeuLeuAspValAlaLeuGlnGlnSerAlaThr 809
                 :::|||      ||||||   ||||||         ||||||:::||||||||||||
Db       2415 GCAATTGATGTGGAAGAGGCAAAACGCCTCCACCGGGAGGCTCTGAAGCAGTCTGCAACT 2474

Qy        810 AspHisAlaThrGlyThrIleAspMetAspLeuIleThrThrGlyValSerAlaSerGlu 829
              |||      ||||||   :::|||:::   ::::::|||||||||:::|||||||:::
Db       2475 GACCCTCGTACTGGCATTGTGGATATTTCTATTCTTACTACAGGAATGAGTGCCACTTCT 2534

Qy        830 ArgIleArgArgAlaAsnLeuLeuAlaAlaLeuArgGluLeuIleAlaAspLysIleSer 849
              |||   |||:::      |||      |||||||||:::||||||         :::|||
Db       2535 CGTAAACGGAAAGAAGAATTAGCTGAAGCATTGAGAAAACTTATT---------TTATCT 2585

Qy        850 ProGlySerSerSerGlyLeuLysThrSerGlnLeuLeuGluAspIleArgSerGlnSer 869
                 |||   :::      ||||||      ||||||   ||||||||||||   ||||||
Db       2586 AAGGGTAAAACACCAGCCTTAAAGTACCAACAGCTGTTTGAGGATATTCGGGGACAGTCT 2645

Qy        870 SerValAspValSerLeuGlnAspIleLysAsnAlaLeuGlySerLeuGlnGlyGluGly 889
                          ::::::      :::   ||||||   :::|||         :::
Db       2646 GACACAGCAATTACCAAGGACATGTTTGAAGAAGCCCTGCGAGCTTTGGCTGATGATGAT 2705

Qy        890 PheLeuThrValHisGlyAspIleValLys 899
              ||||||||||||   |||      |||:::
Db       2706 TTCCTAACAGTGACTGGAAGACTGTCCGC 2735
```

```
RESULT 7
AY408033
LOCUS       AY408033                2522 bp    DNA     linear   GSS 12-DEC-2003
DEFINITION  Homo sapiens HCM3089 gene, VIRTUAL TRANSCRIPT, partial sequence,
            genomic survey sequence.
ACCESSION   AY408033
VERSION     AY408033.1  GI:39764004
KEYWORDS    GSS.
SOURCE      Homo sapiens (human)
  ORGANISM  Homo sapiens
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Euarchontoglires; Primates; Catarrhini;
            Hominidae; Homo.
REFERENCE   1  (bases 1 to 2522)
  AUTHORS   Clark,A.G., Glanowski,S., Nielson,R., Thomas,P., Kejariwal,A.,
            Todd,M.A., Tanenbaum,D.M., Civello,D.R., Lu,F., Murphy,B.,
            Ferriera,S., Wang,G., Zheng,X.H., White,T.J., Sninsky,J.J.,
            Adams,M.D. and Cargill,M.
  TITLE     Inferring nonneutral evolution from human-chimp-mouse orthologous
            gene trios
  JOURNAL   Science 302 (5652), 1960-1963 (2003)
   PUBMED   14671302
REFERENCE   2  (bases 1 to 2522)
  AUTHORS   Clark,A.G., Glanowski,S., Nielson,R., Thomas,P., Kejariwal,A.,
            Todd,M.A., Tanenbaum,D.M., Civello,D.R., Lu,F., Murphy,B.,
            Ferriera,S., Wang,G., Zheng,X.H., White,T.J., Sninsky,J.J.,
            Adams,M.D. and Cargill,M.
  TITLE     Direct Submission
  JOURNAL   Submitted (16-NOV-2003) Celera Genomics, 45 West Gude Drive,
            Rockville, MD 20850, USA
COMMENT     This sequence was made by sequencing genomic exons and ordering
            them based on alignment.
FEATURES             Location/Qualifiers
     source          1..2522
                     /organism="Homo sapiens"
                     /mol_type="genomic DNA"
                     /db_xref="taxon:9606"
```

```
     gene            <1. .>2522
                     /locus_tag="HCM3089"
ORIGIN

Alignment Scores:
Pred. No.:              1.11e-144      Length:         2522
Score:                  1803.50        Matches:        352
Percent Similarity:     62.67%         Conservative:   157
Best Local Similarity:  44.75%         Mismatches:     258
Query Match:            39.13%         Indels:         69
DB:                     10             Gaps:           16

US-10-768-511-8 (1-901) x AY408033 (1-2522)

Qy     49 ArgGlnThrProThrSerAlaValArgArgArgGlyArgGluThrAspSerAlaArgArg 68
Db     18 AGGTCATCTCCCTCTCAG------AGACGTAGAGGCGAGGATTCCACCTCCACGGGGGAG 71

Qy     69 ArgArgSerArgSerArgSerLeuGly--------------------------AsnSer 79
Db     72 TTGCAGCCGATGCCAACCTCGCCTGGAGTGGACCTGCAGAGCCCTGCTGCGCAGGACGTG 131

Qy     80 ValTyrSerSerProTyrAspAlaGlyThrProGlyThrPro---GlyThrProValAla 98
Db    132 CTGTTTTCCAGCCCTCCCCAAATGCATTCTTCAGCTATCCCTCTTGACTTTGATGTTAGT 191

Qy     99 ThrProVal---TyrAlaThrProValGlyThrProMetGlyThrPro---SerPheHis 116
Db    192 TCACCACTGACATACGGCACTCCCAGCTCTCGGGTAGAGGGAACCCCAAGAAGTGGTGTT 251

Qy    117 ArgGlyThrProGlnTyrLysGlnArgSerGluLeuGlySerGlnGlyLysProLeuHis 136
Db    252 AGGGGCACACCT---GTGAGACAGAGGCCTGACCTGGGCTCTGCACAGAAGGGCCTGCAA 308

Qy    137 ArgArgArgArgSerGlnSerArgGluProGlyHisArgSerProSerArgGluProSer 156
Db    309 GTGGATCTGCAGTCT--------------------------------------------- 323

Qy    157 AlaAspGlyArgProSerGluSerAlaGluProAspAspThrLeuGlyGlyGluTyrAla 176
Db    324 ---GACGGGGCAGCAGCAGAAGATATAGTGGCAAGTGAGCAGTCTCTAGGCCAAAAACTT 380

Qy    177 TyrValTrpGlyThrAsnValAsnIleProAspValLeuArgAlaIleArgArgPheLeu 196
Db    381 GTGATCTGGGGAACAGATGTAAATGTGGCAGCATGCAAAGAAAACTTTCAGAGATTTCTT 440

Qy    197 HisAsnTyr-----------------------ArgSerSerAlaHisAspLeuAsnSer 208
Db    441 CAGCGTTTTATTGACCCTCTGGCTAAAGAAGAAGAAAATGTTGGCATAGATATTACTGAA 500

Qy    209 Lys---TyrIleGlnIleIleGluGluThrValGluArgGluGluAspThrLeuAsnIle 227
Db    501 CCTCTATACATGCAACGACTTGGGGAGATTAATGTTATTGGTGAGCCATTTTTAAATGTG 560

Qy    228 AspMetSerAspIleTyrAspHisAspProAspLeuTyrAlaLysIleValArgTyrPro 247
Db    561 AACTGTGAACACATCAAATCATTTGACAAAAATTTGTACAGACAACTCATCTCTTACCCA 620

Qy    248 LeuAspIleIleProLeuLeuAspThrGluCysGlnGluValAlaThrSerLeuLeuPro 267
Db    621 CAGGAAGTTATTCCAACTTTTGACATGGCTGTCAATGAAATCTCCTTTGACCGTTACCCT 680

Qy    268 ------ThrPheGluLysHisIleGluAlaArgProPheAsnLeuLysAlaSerValHis 285
Db    681 GACTCAATCTTAGAACATCAGATTCAAGTAAGACCATTCAACGCATTGAAGACTAAGAAT 740

Qy    286 MetArgGluLeuAsnProSerAspIleAspLysLeuValSerValLysGlyMetValIle 305
Db    741 ATGAGAAACCTGAATCCAGAAGACATTGACCAGCTCATCACCATCAGCGGCATGGTGATC 800

Qy    306 ArgCysSerSerIleIleProGluIleLysGlyAlaPhePheLysCysLeuValCysGly 325
Db    801 AGGACATCCCAGCTGATTCCCGAGATGCAGGAGGCCTTCTTCCAGTGCCAAGTGTGTGCC 860

Qy    326 HisSerProProLeuValThrValValLysGlyArgValGluGluProThrArgCysGlu 345
Db    861 CACACG---ACCCGGGTGGAGATGGACCGCGGCCGCATTGCAGAGCCCAGTGTGTGCGGG 917

Qy    346 LysProGluCysAlaAlaArgAsnAlaMetSerLeuIleHisAsnArgCysThrPheAla 365
Db    918 CGC------TGCCACACCACCCACAGCATGGCACTCATCCACAACCGCTCCCTCTTCTCT 971

Qy    366 AsnLysGlnIleValArgLeuGlnGluThrProAspAlaIleProGluGlyGluThrPro 385
Db    972 GACAAGCAGATGATCAAGCTTCAGGAGTCTCCGGAAGACATGCCTGCAGGGCAGACACCA 1031

Qy    386 HisThrValSerMetCysLeuTyrAsnThrMetValAspAlaValLysProGlyAspArg 405
Db   1032 CACACAGTTATCCTGTTTGCTCACAATGATCTCGTTGACAAGGTCCAGCCTGGGGACAGA 1091

Qy    406 IleGluValThrGlyValPheLysAlaMetAlaValArgValGlyProAsnGlnArgThr 425
Db   1092 GTGAATGTTACAGGCATCTATCGAGCTGTGCCTATTCGAGTCAATCCAAGAGTGAGTAAT 1151

Qy    426 LeuArgAlaLeuTyrLysThrTyrIleAspCysValHisValLysLysSerAspArgGly 445
Db   1152 GTGAAGTCTGTCTACAAAACCCACATTGATGTCATTCATTATCGGAAAACGGATGCAAAA 1211

Qy    446 ArgLeuGlnThrGluAspProMetGluMetAspLysGluAsnAspMetTyrAlaGlyTyr 465
Db   1212 CGTCTGCAT------------------------------------------GGCCTT 1226

Qy    466 HisGluSerAspThrSerGluAlaAlaAsnGluAlaLysIleGlnLysLeuLysGluLeu 485
Db   1227 GATGAAGAAGCAGAACAGAAACTTTTTCAGAGAAACGTGTGGAATTGCTTAAGGAACTT 1286

Qy    486 SerLysLeuProGlyIleTyrAspArgLeuSerArgSerLeuAlaProSerIleTrpGlu 505
Db   1287 TCCAGGAAACCAGACATTTATGAGAGGCTTGCTTCAGCCTTGGCTCCAAGCATTTATGAA 1346

Qy    506 LeuGluAspIleLysLysGlyLeuLeuCysGlnLeuPheGlyGlyLysAlaLysLysIle 525
Db   1347 CATGAAGATATAAAGAAGGGAATTTTGCTTCAGCTCTTTGGCGGGACAAGGAAGGATTTT 1406

Qy    526 ------ProSerGlyAlaSerPheArgGlyAspIleAsnValLeuLeuValGlyAspPro 543
Db   1407 AGTCACACTGGAAGGGGCAAATTTCGGGCTGAGATCAACATCTTGCTGTGTGGCGACCCT 1466

Qy    544 GlyThrSerLysSerGlnLeuLeuGlnTyrValHisLysIleAlaProArgGlyIleTyr 563
Db   1467 GGTACCAGCAAGTCCCAGCTGCTGCAGTACGTGTACAACCTCGTCCCCAGGGGCCAGTAC 1526

Qy    564 ThrSerGlyArgGlySerSerAlaValGlyLeuThrAlaTyrValThrLysAspProGlu 583
Db   1527 ACGTCTGGGAAGGGCTCCAGTGCAGTTGGCCTCACTGCGTACGTAATGAAAGACCCTGAG 1586

Qy    584 ThrArgGluThrValLeuGluSerGlyAlaLeuValLeuSerAspArgGlyIleCysCys 603
Db   1587 ACAAGGCAGCTGGTCCTGCAGACAGGTGCTCTTGTCCTGAGTGACAACGGCATCTGCTGT 1646

Qy    604 IleAspGluPheAspLysMetSerAspAsnAlaArgSerMetLeuHisGluValMetGlu 623
Db   1647 ATCGATGAGTTCGACAAGATGAATGAAAGTACAAGATCGGTATTGCATGAAGTCATGGAA 1706

Qy    624 GlnGlnThrValSerValAlaLysGlyGlyIleIleAlaSerLeuAsnAlaArgThrSer 643
Db   1707 CAGCAGACTCTGTCCATTGCAAAGGCTGGGATCATCTGTCAGCTCAATGCGCGCACCTCT 1766

Qy    644 ValLeuAlaCysAlaAsnProSerGlySerArgTyrAsnAlaArgLeuSerValIleAsp 663
Db   1767 GTCCTGGCAGCAGCAAATCCCATTGAGTCTCAGTGGAATCCTAAAAAAACAACCATTGAA 1826

Qy    664 AsnIleGlnLeuProProThrLeuLeuSerArgPheAspLeuIleTyrLeuMetLeuAsp 683
Db   1827 AACATCCAGCTGCCTCATACTTTATTATCAAGGTTTGATTTGATCTTCCTCATGCTGGAC 1886
```

```
Qy        684 LysProAspGluGlnAsnAspArgArgLeuAlaArgHisLeuValAlaLeuHisTyrGlu 703
                 ||||||         ||||||||||||||    ||||||||||||||:::|||:::
Db       1887 CCTCAGGACGAAGCCTATGACAGGCGTCTGGCTCACCACCTGGTCGCACTGTACTACCAG 1946

Qy        704 AsnTyrGluValSerLysGlnAspAlaLeuAspLeuGlnThrLeuThrAlaTyrIleThr 723
              :::  |||  :::::::::::: ||||||:::       |||      ||||||
Db       1947 AGCGAGGAGCAGGCAGAGGAGGAGCTCCTGGACATGGCGGTGCTAAAGGACTACATTGCC 2006

Qy        724 TyrAlaArgGlnHisValHisProThrLeuSerAspGluAlaAlaGluAspLeuIleAsn 743
              ||||||       :::   |||   ||||||:::||||||::::::   ||||||
Db       2007 TACGCGCACAGCACCATCATGCCGCGGCTAAGTGAGGAAGCCAGCCAGGCTCTCATCGAG 2066

Qy        744 GlyTyrValGluMetArgGlnLysGlyAsnPheProGlySerSerLysLysValIleThr 763
                 ||||||:::||||||:::                 |||||||||:::   :::::::::
Db       2067 GCTTATGTAGACATGAGGAAG------------ATTGGCAGTAGCCGGGGAATGGTTTCT 2114

Qy        764 AlaThrProArgGlnLeuGluSerMetIleArgIleSerGluAlaLeuAlaArgMetArg 783
              |||   ||||||||||||||||||:::||||||::::::||||||   |||::::::|||
Db       2115 GCATACCCTCGACAGCTAGAGTCATTAATCCGCTTAGCAGAAGCCCATGCTAAAGTAAGA 2174

Qy        784 PheSerGluValValGluLysValAspAlaAlaGluAlaValArgLeuLeuAspValAla 803
                 |||    ||||||   :::|||      ||||||    ||||||            |||
Db       2175 TTGTCTAACAAAGTTGAAGCCATTGATGTGGAAGAGGCCAAACGCCTCCATCGGGAAGCT 2234

Qy        804 LeuGlnGlnSerAlaThrAspHisAlaThrGlyThrIleAspMetAspLeuIleThrThr 823
              |||:::|||||||||||||||         ||||||   :::|||:::   ::::::||||||
Db       2235 CTGAAGCAGTCTGCAACTGATCCCCGGACTGGCATCGTGGACATATCTATTCTTACTACG 2294

Qy        824 GlyValSerAlaSerGluArgIleArgArgAlaAsnLeuLeuAlaAlaLeuArgGluLeu 843
              |||:::||||||:::   |||   |||:::         |||      ||||||::::::|||
Db       2295 GGGATGAGTGCCACCTCTCGTAAACGGAAAGAAGAATTAGCTGAAGCATTGAAAAAGCTT 2354

Qy        844 IleAlaAspLysIleSerProGlySerSerSerGlyLeuLysThrSerGlnLeuLeuGlu 863
              |||      :::|||   |||   :::         ||||||      ||||||   |||
Db       2355 ATT---------TTATCTAAGGGCAAAACACCAGCTCTAAAATACCAGCAACTTTTTGAA 2405

Qy        864 AspIleArgSerGlnSerSerValAspValSerLeuGlnAspIleLysAsnAlaLeuGly 883
              |||||||||   ||||||      :::   ::::::            :::   ||||||
Db       2406 GATATTCGGGGACAATCTGACATAGCAATTACTAAAGATATGTTGAAGAAGCACTGCGT 2465

Qy        884 SerLeuGlnGlyGluGlyPheLeuThrValHisGlyAspIleValLys 899
              :::|||       :::      ||||||||||||   |||      |||:::
Db       2466 GCCCTGGCAGATGATGATTTCCTGACAGTGACTGGGAAGACCGTGCGC 2513
```

```
RESULT 8
AY408035
LOCUS       AY408035           2519 bp    DNA     linear   GSS 12-DEC-2003
DEFINITION  Mus musculus HCM3089 gene, VIRTUAL TRANSCRIPT, partial sequence,
            genomic survey sequence.
ACCESSION   AY408035
VERSION     AY408035.1  GI:39764006
KEYWORDS    GSS.
SOURCE      Mus musculus (house mouse)
  ORGANISM  Mus musculus
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Euarchontoglires; Glires; Rodentia;
            Sciurognathi; Muroidea; Muridae; Murinae; Mus.
REFERENCE   1  (bases 1 to 2519)
  AUTHORS   Clark,A.G., Glanowski,S., Nielson,R., Thomas,P., Kejariwal,A.,
            Todd,M.A., Tanenbaum,D.M., Civello,D.R., Lu,F., Murphy,B.,
            Ferriera,S., Wang,G., Zheng,X.H., White,T.J., Sninsky,J.J.,
            Adams,M.D. and Cargill,M.
  TITLE     Inferring nonneutral evolution from human-chimp-mouse orthologous
            gene trios
  JOURNAL   Science 302 (5652), 1960-1963 (2003)
   PUBMED   14671302
REFERENCE   2  (bases 1 to 2519)
  AUTHORS   Clark,A.G., Glanowski,S., Nielson,R., Thomas,P., Kejariwal,A.,
            Todd,M.A., Tanenbaum,D.M., Civello,D.R., Lu,F., Murphy,B.,
            Ferriera,S., Wang,G., Zheng,X.H., White,T.J., Sninsky,J.J.,
            Adams,M.D. and Cargill,M.
  TITLE     Direct Submission
  JOURNAL   Submitted (16-NOV-2003) Celera Genomics, 45 West Gude Drive,
            Rockville, MD 20850, USA
COMMENT     This sequence was made by sequencing genomic exons and ordering
            them based on alignment.
FEATURES             Location/Qualifiers
     source          1..2519
                     /organism="Mus musculus"
                     /mol_type="genomic DNA"
                     /db_xref="taxon:10090"
     gene            <1..>2519
                     /locus_tag="HCM3089"
ORIGIN

Alignment Scores:
Pred. No.:              2.46e-144       Length:         2519
Score:                  1799.50         Matches:        393
Percent Similarity:     60.88%          Conservative:   147
Best Local Similarity:  44.31%          Mismatches:     250
Query Match:            39.04%          Indels:         97
DB:                     10              Gaps:           16

US-10-768-511-8 (1-901) x AY408035 (1-2519)
```

```
Qy         59 ArgGlyArgGluThrAspSerAla------ArgArgArgArgSerArgSerArgSerLeu 76
              |||      |||:::   |||:::          ||||||||||||            |||
Db          3 CGAAGTGAGGAAAGCAGGTCGTCACCCAATCGGAGACGTAGAGGCGAAGATTCTTCCACC 62

Qy         77 Gly--------------------------------------------------------- 77
              |||
Db         63 GGAGAGCTACTGCCAATGCCCACCTCACCAGGAGCCGACCTGCAGAGCCCACCTGCACAG 122

Qy         78 AsnSerValTyrSerSerPro-----------------------------------TyrAsp 86
              |||::::::::::|||||||||                                  :::|||
Db        123 AATGCCTTGTTTTCCAGCCCTCCTCAGATGCATTCTTTAGCTATTCCTTTGGACTTTGAT 182

Qy         87 AlaGlyThrProGlyThrProGlyThrProValAlaThrProValTyrAlaThrProVal 106
                 :::|||   |||                              |||   ||||||
Db        183 GTTAGTTCACCATTGACA--------------------------TATGGCACTCCCAGC 215

Qy        107 GlyThrProMetGlyThrPro---SerPheHisArgGlyThrProGlnTyrLysGlnArg 125
                       |||||||||   |||      ||||||||||||         :::||||||
Db        216 TCGCGAGTGGAAGGAACCCCAAGAAGTGGGGTGAGAGGCACACCT---GTAAGGCAGAGG 272

Qy        126 SerGluLeuGlySerGlnGlyLysProLeuHisArgArgArgArgSerGlnSerArgGlu 145
                 :::||||||||||      |||   |||            :::|||
Db        273 CCAGATCTGGGCTCAGCACGAAAGGGTTTGCAGGTGGATCTGCAGTCT------------ 320

Qy        146 ProGlyHisArgSerProSerArgGluProSerAlaAspGlyArgProSerGluSerAla 165
                                                  ||||||      :::|||
Db        321 -----------------------------------GATGGCGCAGCAGCAGAAGACATC 344

Qy        166 GluProAspAspThrLeuGlyGlyGluTyrAlaTyrValTrpGlyThrAsnValAsnIle 185
                 |||   :::         |||:::         :::|||||||||:::|||||||:::
Db        345 GTACCAAGTGAACAGTCTCTAGGCCAAAAGCTTGTGATTTGGGGAACAGATGTGAATGTG 404

Qy        186 ProAspValLeuArgAlaIleArgArgPheLeuHisAsnTyr------------------ 199
                                 :::|||||||||      :::
Db        405 GCAACATGTAAAGAGAATTTCAGAGATTCCTTCAGTGTTTTACTGATCCTCTGGCCAAA 464

Qy        200 ------ArgSerSerAlaHisAspLeuAsnSerLys---TyrIleGlnIleIleGluGlu 216
                    :::         |||:::                |||:::|||   :::   |||
Db        465 GAAGAAGAAAATGTTGGCATAGATATTACTCAACCTTTGTACATGCAACAACTTGGAGAG 524

Qy        217 ThrValGluArgGluGluAspThrLeuAsnIleAspMetSerAspIleTyrAspHisAsp 236
                          |||      ||||||::::::                  |||
Db        525 ATTAATATTACAGGAGAGCCATTTTAAATGTGAACTGCGAACACATAAAATCATTTAGC 584

Qy        237 ProAspLeuTyrAlaLysIleValArgTyrProLeuAspIleIleProLeuLeuAspThr 256
                 :::||||||   :::::::::   ||||||   ::::::||||||      |||
Db        585 AAAAATCTGTATAGACAGCTCATCTCCTACCCACAGGAGGTTATACCAACCTTTGACATG 644

Qy        257 GluCysGlnGluValAlaThrSerLeuLeuPro------ThrPheGluLysHisIleGlu 274
                    |||:::               |||            |||      |||:::
Db        645 GCTGTCAATGAGATCTTCTTTGACCGTTATCCTGACTCCATCTTAGAACATCAGATTCAA 704
```

```
Qy    275 AlaArgProPheAsnLeuLysAlaSerValHisMetArgGluLeuAsnProSerAspIle 294
Db    705 GTCAGACCTTTTAATGCGTTGAAGACAAAGAGTATGAGAAACTTGAATCCAGAAGACATT 764

Qy    295 AspLysLeuValSerValLysGlyMetValIleArgCysSerSerIleIleProGluIle 314
Db    765 GATCAGCTCATCACCATCAGTGGCATGGTCATCAGAACATCACAGCTGATTCCGGAGATG 824

Qy    315 LysGlyAlaPhePheLysCysLeuValCysGlyHisSerProProLeuValThrValVal 334
Db    825 CAGGAGGCCTTTTTCCAATGCCAAGTCTGTGCCCACACC---ACCCGGGTGGAGATAGAT 881

Qy    335 LysGlyArgValGluGluProThrArgCysGluLysProGluCysAlaAlaArgAsnAla 354
Db    882 CGAGGCAGAATTGCTGAGCCCTGCAGTTGT------GTGCACTGCCACACTACCCACAGC 935

Qy    355 MetSerLeuIleHisAsnArgCysThrPheAlaAsnLysGlnIleValArgLeuGlnGlu 374
Db    936 ATGGCACTGATCCACAACCGATCATTCTTCTCTGACAAGCAAATGATCAAACTTCAAGAG 995

Qy    375 ThrProAspAlaIleProGluGlyGluThrProHisThrValSerMetCysLeuTyrAsn 394
Db    996 TCTCCTGAAGACATGCCTGCTGGGCAGACACCTCACACTATTGTCCTTTTGCCCACAAT 1055

Qy    395 ThrMetValAspAlaValLysProGlyAspArgIleGluValThrGlyValPheLysAla 414
Db   1056 GACCTTGTTGACAAGGTTCAACCAGGGGACAGAGTGAACGTCACAGGCATATATCGAGCA 1115

Qy    415 MetAlaValArgValGlyProAsnGlnArgThrLeuArgAlaLeuTyrLysThrTyrIle 434
Db   1116 GTACCAATTCGAGTTAATCCAAGAGTGAGCAACGTGAAGTCTGTCTATAAAACCCACATT 1175

Qy    435 AspCysValHisValLysLysSerAspArgGlyArgLeuGlnThrGluAspProMetGlu 454
Db   1176 GATGTCATTCATTATCGGAAAACGGATGCAAAACGTCTGCAT------------------ 1217

Qy    455 MetAspLysGluAsnAspMetTyrAlaGlyTyrHisGluSerAspThrSerGluAlaAla 474
Db   1218 ---------------------------GGCCTTGATGAAGAAGCAGAACAGAAACTTTTT 1250

Qy    475 AsnGluAlaLysIleGlnLysLeuLysGluLeuSerLysLeuProGlyIleTyrAspArg 494
Db   1251 TCAGAGAAACGTGTGAAATTGCTTAAGGAACTTTCCAGGAAGCCAGATATTTATGAGCGG 1310

Qy    495 LeuSerArgSerLeuAlaProSerIleTrpGluLeuGluAspIleLysLysGlyLeuLeu 514
Db   1311 CTTGCTTCAGCCTTGGCTCCCAGCATTTATGAACATGAAGATATCAAAAAGGGAATCTTA 1370

Qy    515 CysGlnLeuPheGlyGlyLysAlaLysLysIle------ProSerGlyAlaSerPheArg 532
Db   1371 CTTCAGCTCTTTGGTGGAACAAGGAAGGATTTCAGTCACACTGGAGGGGTAAATTCCGT 1430

Qy    533 GlyAspIleAsnValLeuLeuValGlyAspProGlyThrSerLysSerGlnLeuLeuGln 552
Db   1431 GCTGAGATCAACATCCTTCTGTGTGGGGACCCTGGCACCAGCAAGTCCCAGCTGCTACAG 1490

Qy    553 TyrValHisLysIleAlaProArgGlyIleTyrThrSerGlyArgGlySerSerAlaVal 572
Db   1491 TATGTGTACAACCTGGTCCCCAGAGGCCAGTACACGTCTGGAAAAGGCTCCAGTGCGGTC 1550

Qy    573 GlyLeuThrAlaTyrValThrLysAspProGluThrArgGluThrValLeuGluSerGly 592
Db   1551 GGCCTCACCGCCTATGTGATGAAAGACCCTGAGACCAGGCAGCTTGTCCTCCAGACAGGT 1610

Qy    593 AlaLeuValLeuSerAspArgGlyIleCysCysIleAspGluPheAspLysMetSerAsp 612
Db   1611 GCCCTCGTCCTGAGTGACAATGGGATATGCTGCATCGATGAGTTTGACAAAATGAATGAA 1670

Qy    613 AsnAlaArgSerMetLeuHisGluValMetGluGlnGlnThrValSerValAlaLysGly 632
Db   1671 AGCACAAGGTCTGTGCTGCATGAGGTCATGGAACAGCAGACTCTGTCCATTGCAAAGGCT 1730

Qy    633 GlyIleIleAlaSerLeuAsnAlaArgThrSerValLeuAlaCysAlaAsnProSerGly 652
Db   1731 GGGATCATCTGTCAGCTCAATGCGCGCACCTCTGTCCTGGCAGCAGCAAATCCTATTGAG 1790

Qy    653 SerArgTyrAsnAlaArgLeuSerValIleAspAsnIleGlnLeuProProThrLeuLeu 672
Db   1791 TCTCAGTGGAATCCTAAAAAAACAACCATTGAAAATATCCAACTACCGCACACATTGTTG 1850

Qy    673 SerArgPheAspLeuIleTyrLeuMetLeuAspLysProAspGluGlnAsnAspArgArg 692
Db   1851 TCAAGGTTTGATCTCATTTTCCTCATGCTAGACCCTCAGGATGAGGCATATGACCGGCGT 1910

Qy    693 LeuAlaArgHisLeuValAlaLeuHisTyrGluAsnTyrGluValSerLysGlnAspAla 712
Db   1911 CTAGCTCATCACCTGGTTTCATTGTACTACCAAAGTGAGGAGCAAGTGGAGGAGGAGTTC 1970

Qy    713 LeuAspLeuGlnThrLeuThrAlaTyrIleThrTyrAlaArgGlnHisValHisProThr 732
Db   1971 CTGGACATGGCCGTGCTGAAAGACTACATTGCATATGCCCATAGTACCATCATGCCCCGA 2030

Qy    733 LeuSerAspGluAlaAlaGluAspLeuIleAsnGlyTyrValGluMetArgGlnLysGly 752
Db   2031 CTGAGTGAGGAGGCCAGCCAGGCTCTCATTGAGGCTTATGTAAACATGAGGAAG------ 2084

Qy    753 AsnPheProGlySerSerLysLysValIleThrAlaThrProArgGlnLeuGluSerMet 772
Db   2085 ------ATTGGGAGTAGCCGGGGGATGGTTTCTGCTTACCCTCGACAGCTAGAGTCATTA 2138

Qy    773 IleArgIleSerGluAlaLeuAlaArgMetArgPheSerGluValValGluLysValAsp 792
Db   2139 ATTCGCTTAGCAGAAGCCCATGCTAAAGTAAGATTTTCAAACAAAGTTGAAGCAATTGAT 2198

Qy    793 AlaAlaGluAlaValArgLeuLeuAspValAlaLeuGlnGlnSerAlaThrAspHisAla 812
Db   2199 GTGGAAGAGGCAAAACGCCTCCACCGGGAGGCTCTGAAGCAGTCTGCAACTGACCCTCGT 2258

Qy    813 ThrGlyThrIleAspMetAspLeuIleThrThrGlyValSerAlaSerGluArgIleArg 832
Db   2259 ACTGGCATTGTGGATATTTCTATTCTTACTACAGGAATGAGTGCCACTTCTCGTAAACGG 2318

Qy    833 ArgAlaAsnLeuLeuAlaAlaLeuArgGluLeuIleAlaAspLysIleSerProGlySer 852
Db   2319 AAAGAAGAATTAGCTGAAGCATTGAGAAAACTTATT---------TTATCTAAGGGTAAA 2369

Qy    853 SerSerGlyLeuLysThrSerGlnLeuLeuGluAspIleArgSerGlnSerSerValAsp 872
Db   2370 ACACCAGCCTTAAAGTACCAACAGCTGTTGAGGATATTCGGGGACAGTCTGACACAGCA 2429

Qy    873 ValSerLeuGlnAspIleLysAsnAlaLeuGlySerLeuGlnGlyGluGlyPheLeuThr 892
Db   2430 ATTACCAAGGACATGTTTGAAGAAGCCCTGCGAGCTTTGGCTGATGATGATTTCCTAACA 2489

Qy    893 ValHisGlyAspIleValLys 899
Db   2490 GTGACTGGGAAGACTGTCCGC 2510
```

RESULT 9
AY408034
LOCUS   AY408034   2357 bp   DNA   linear   GSS 12-DEC-2003
DEFINITION Pan troglodytes HCM3089 gene, VIRTUAL TRANSCRIPT, partial sequence, genomic survey sequence.
ACCESSION   AY408034
VERSION   AY408034.1 GI:39764005
KEYWORDS   GSS.
SOURCE   Pan troglodytes (chimpanzee)
  ORGANISM Pan troglodytes
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Euarchontoglires; Primates; Catarrhini; Hominidae; Pan.
REFERENCE   1 (bases 1 to 2357)
  AUTHORS   Clark,A.G., Glanowski,S., Nielson,R., Thomas,P., Kejariwal,A., Todd,M.A., Tanenbaum,D.M., Civello,D.R., Lu,F., Murphy,B., Ferriera,S., Wang,G., Zheng,X.H., White,T.J., Sninsky,J.J.,

```
            Adams,M.D. and Cargill,M.
  TITLE     Inferring nonneutral evolution from human-chimp-mouse orthologous
            gene trios
  JOURNAL   Science 302 (5652), 1960-1963 (2003)
   PUBMED   14671302
REFERENCE   2  (bases 1 to 2357)
  AUTHORS   Clark,A.G., Glanowski,S., Nielson,R., Thomas,P., Kejariwal,A.,
            Todd,M.A., Tanenbaum,D.M., Civello,D.R., Lu,F., Murphy,B.,
            Ferriera,S., Wang,G., Zheng,X.H., White,T.J., Sninsky,J.J.,
            Adams,M.D. and Cargill,M.
  TITLE     Direct Submission
  JOURNAL   Submitted (16-NOV-2003) Celera Genomics, 45 West Gude Drive,
            Rockville, MD 20850, USA
COMMENT     This sequence was made by sequencing genomic exons and ordering
            them based on alignment.
FEATURES             Location/Qualifiers
     source          1..2357
                     /organism="Pan troglodytes"
                     /mol_type="genomic DNA"
                     /db_xref="taxon:9598"
     gene            <1..>2357
                     /locus_tag="HCM3089"
ORIGIN

Alignment Scores:
Pred. No.:              2.14e-132     Length:         2357
Score:                  1661.00       Matches:        349
Percent Similarity:     64.01%        Conservative:   133
Best Local Similarity:  46.35%        Mismatches:     209
Query Match:            36.04%        Indels:         62
DB:                     10            Gaps:           12

US-10-768-511-8 (1-901) x AY408034 (1-2357)

Qy         85 TyrAspAlaGlyThrProGlyThrProGlyThrProValAlaThrProValTyrAlaThr 104
Db         15 TTTGATGTTAGTTCACCACTGACA--------------------------TACGGCACT 47

Qy        105 ProValGlyThrProMetGlyThrPro---SerPheHisArgGlyThrProGlnTyrLys 123
Db         48 CCCAGCTCTCGGGTAGAGGGAACCCCAAGAAGTGGTGTTAGGGGCACACCT---GTGAGA 104

Qy        124 GlnArgSerGluLeuGlySerGlnGlyLysProLeuHisArgArgArgArgSerGlnSer 143
Db        105 CAGAGGCCTGACCTGGGCTCTGCACAGAAGGGCCTGCAAGTGGATCTGCAGTCT------ 158

Qy        144 ArgGluProGlyHisArgSerProSerArgGluProSerAlaAspGlyArgProSerGlu 163
Db        159 -----------------------------------------GACGGGGCAGCAGCAGAA 176

Qy        164 SerAlaGluProAspAspThrLeuGlyGlyGluTyrAlaTyrValTrpGlyThrAsnVal 183
Db        177 GATATAGTGGCAAGCGAGCAGTCTCTAGGCCAAAAACTTGTGATCTGGGGAACAGATGTA 236

Qy        184 AsnIleProAspValLeuArgAlaIleArgArgPheLeuHisAsnTyr------------ 199
Db        237 AATGTGGCAGCATGCAAAGAAAACTTTCAGAGATTTCTTCAGCGTTTTATTGACCCTCTG 296

Qy        200 ------------ArgSerSerAlaHisAspLeuAsnSerLys---TyrIleGlnIleIle 214
Db        297 GCTAAAGAAGAAGAAAATGTTGGCATAGATATTACTGAACCTCTATACATGCAACGACTT 356

Qy        215 GluGluThrValGluArgGluGluAspThrLeuAsnIleAspMetSerAspIleTyrAsp 234
Db        357 GGGGAGATTAATGTTATTGGTGAGCCATTTTTAAATGTGAACTGTGAACACATCAAATCA 416

Qy        235 HisAspProAspLeuTyrAlaLysIleValArgTyrProLeuAspIleIleProLeuLeu 254
Db        417 TTTGACAAAAAATTTGTACAGACAACTCATCTCTTACCCACAGGAAGTTATTCCAACTTTT 476

Qy        255 AspThrGluCysGlnGluValAlaThrSerLeuLeuPro------ThrPheGluLysHis 272
Db        477 GACATGGCTGTCAATGAAATCTTCTTTGACCGTTACCCTGACTCAATCTTAGAACATCAG 536

Qy        273 IleGluAlaArgProPheAsnLeuLysAlaSerValHisMetArgGluLeuAsnProSer 292
Db        537 ATTCAAGTAAGACCATTCAACGCATTGAAGACTAAGAATATGAGAAACCTGAATCCAGAA 596

Qy        293 AspIleAspLysLeuValSerValLysGlyMetValIleArgCysSerSerIleIlePro 312
Db        597 GACATTGACCAGCTCATCACCATCAGTGGCATGGTGATCAGGACATCCCAGCTGATTCCC 656

Qy        313 GluIleLysGlyAlaPhePheLysCysLeuValCysGlyHisSerProProLeuValThr 332
Db        657 GAGATGCAGGAGGCCTTCTTCCAGTGCCAAGTGTGTGCCCACACG---ACCCGGGTGGAG 713

Qy        333 ValValLysGlyArgValGluGluProThrArgCysGluLysProGluCysAlaAlaArg 352
Db        714 ATGGACCGCGGCCGCATTGCAGAGCCCAGTGTGTGCGGGCGC------TGCCACACCACC 767

Qy        353 AsnAlaMetSerLeuIleHisAsnArgCysThrPheAlaAsnLysGlnIleValArgLeu 372
Db        768 CACAGCATGGCACTCATCCACAACCGCTCCCTCTTCTCTGACAAGCAGATGATCAAGCTT 827

Qy        373 GlnGluThrProAspAlaIleProGluGlyGluThrProHisThrValSerMetCysLeu 392
Db        828 CAGGAGTCTCCGGAAGACATGCCTGCAGGGCAGACACCACACACAGTTATCCTGTTTGCT 887

Qy        393 TyrAsnThrMetValAspAlaValLysProGlyAspArgIleGluValThrGlyValPhe 412
Db        888 CACAATGATCTCGTTGACAAGGTCCAGCCTGGGGACAGAGTGAATGTTACAGGCATCTAT 947

Qy        413 LysAlaMetAlaValArgValGlyProAsnGlnArgThrLeuArgAlaLeuTyrLysThr 432
Db        948 CGAGCTGTGCCTATTCGAGTCAATCCAAGAGTGAGTAATGTGAAGTCTGTCTACAAAACC 1007

Qy        433 TyrIleAspCysValHisValLysLysSerAspArgGlyArgLeuGlnThrGluAspPro 452
Db       1008 CACATTGATGTCATTCATTATCGGAAAACGGATGCAAAACGTCTGCAT------------ 1055

Qy        453 MetGluMetAspLysGluAsnAspMetTyrAlaGlyTyrHisGluSerAspThrSerGlu 472
Db       1056 ---------------------------------GGCCTTGATGAAGAAGCAGAACAGAAA 1082

Qy        473 AlaAlaAsnGluAlaLysIleGlnLysLeuLysGluLeuSerLysLeuProGlyIleTyr 492
Db       1083 CTTTTTTCAGAGAAACGTGTGGAATTGCTTAAGGAACTTTCCAGGAAACCAGACATTTAT 1142

Qy        493 AspArgLeuSerArgSerLeuAlaProSerIleTrpGluLeuGluAspIleLysLysGly 512
Db       1143 GAGAGGCTTGCTTCAGCCTTGGCTCCAAGCATTTATGAACATGAAGATATAAAGAAGGGA 1202

Qy        513 LeuLeuCysGlnLeuPheGlyGlyLysAlaLysLysIle------ProSerGlyAlaSer 530
Db       1203 ATTTTGCTTCAGCTCTTTGGCGGGACAAGGAAGGATTTAGTCACACTGGAAGGGGCAAA 1262

Qy        531 PheArgGlyAspIleAsnValLeuLeuValGlyAspProGlyThrSerLysSerGlnLeu 550
Db       1263 TTTCGGGCTGAGATCAACATCTTGCTGTGTGGCGACCCTGGTACCAGCAAGTCCCAGCTG 1322

Qy        551 LeuGlnTyrValHisLysIleAlaProArgGlyIleTyrThrSerGlyArgGlySerSer 570
Db       1323 CTGCAGTATGTGTACAACCTCGTCCCCAGGGGCCAGTACACGTCTGGGAAGGGCTCCAGT 1382

Qy        571 AlaValGlyLeuThrAlaTyrValThrLysAspProGluThrArgGluThrValLeuGlu 590
Db       1383 GCAGTTGGCCTCACTGCGTACGTAATGAAAGACCCTGAGACAAGGCAGCTGGTCCTGCAG 1442

Qy        591 SerGlyAlaLeuValLeuSerAspArgGlyIleCysCysIleAspGluPheAspLysMet 610
Db       1443 ACAGGTGCTCTTGTCCTGAGTGACAACGGCATCTGCTGTATCGATGAGTTCGACAAGATG 1502

Qy        611 SerAspAsnAlaArgSerMetLeuHisGluValMetGluGlnGlnThrValSerValAla 630
Db       1503 AATGAAAGTACAAGATCGGTATTGCATGAAGTCATGGAACAGCAGACTCTGTCCATTGCA 1562
```

```
Qy        631 LysGlyGlyIleIleAlaSerLeuAsnAlaArgThrSerValLeuAlaCysAlaAsnPro 650
              |||   ||||||||     ||||||||||||||||||| ||||||    |||||||||
Db       1563 AAGGCTGGGATCATCTGTCAGCTCAATGCGCGCACCTCTGTCCTGGCAGCAGCAAATCCC 1622

Qy        651 SerGlySerArgTyrAsnAlaArgLeuSerValIleAspAsnIleGlnLeuProProThr 670
                 |||::::::|||   :::   :::   |||:::|||||||||||||||    |||
Db       1623 ATTGAGTCTCAGTGGAATCCTAAAAAAACAACCATTGAAAACATCCAGCTGCCTCATACT 1682

Qy        671 LeuLeuSerArgPheAspLeuIleTyrLeuMetLeuAspLysProAspGluGlnAsnAsp 690
              |||||||||||||||||||||||||:::||||||||||||     ||||||       |||
Db       1683 TTATTATCAAGGTTTGATTTGATCTTCCTCATGCTGGACCCTCAGGACGAAGCCTATGAC 1742

Qy        691 ArgArgLeuAlaArgHisLeuValAlaLeuHisTyrGluAsnTyrGluValSerLysGln 710
              ||||||||||||   |||||||||||||||:::|||::::::   |||    :::::::::
Db       1743 AGGCGTCTGGCTCACCACCTGGTCGCACTGTACTACCAGAGCGAGGAGCAGGCAGAGGAG 1802

Qy        711 AspAlaLeuAspLeuGlnThrLeuThrAlaTyrIleThrTyrAlaArgGlnHisValHis 730
              :::   ||||||:::       |||       ||||||   ||||||          :::
Db       1803 GAGCTCCTGGACATGGCGGTGCTAAAGGACTACATTGCCTACGCACACAGCACCATCATG 1862

Qy        731 ProThrLeuSerAspGluAlaAlaGluAspLeuIleAsnGlyTyrValGluMetArgGln 750
              |||   ||||||:::||||||::::::   ||||||      ||||||:::||||||:::
Db       1863 CCGCGGCTAAGTGAGGAAGCCAGCCAGGCTCTCATCGAGGCTTATGTAGACATGAGGAAG 1922

Qy        751 LysGlyAsnPheProGlySerSerLysLysValIleThrAlaThrProArgGlnLeuGlu 770
                             |||||||||:::   :::::::::|||   ||||||||||||||
Db       1923 ------------ATTGGCAGTAGCCGGGGAATGGTTTCTGCATACCCTCGACAGCTAGAG 1970

Qy        771 SerMetIleArgIleSerGluAlaLeuAlaArgMetArgPheSerGluValValGluLys 790
              |||:::||||||:::::::||||||   |||::::::|||   |||        ||||||
Db       1971 TCATTAATCCGCTTAGCAGAAGCCCATGCTAAAGTAAGATTGTCTAACAAAGTTGAAGCC 2030

Qy        791 ValAspAlaAlaGluAlaValArgLeuLeuAspValAlaLeuGlnGlnSerAlaThrAsp 810
              :::|||      ||||||   ||||||         ||||||:::||||||||||||||
Db       2031 ATTGATGTGGAAGAGGCCAAACGCCTCCANNNNNNNGCTCTGAAGCAGTCTGCAACTGAT 2090

Qy        811 HisAlaThrGlyThrIleAspMetAspLeuIleThrThr 823
                       ||||||   :::|||:::   ::::::||||||
Db       2091 CCCCGGACTGGCATCGTGGACATATCTATTCTTACTACG 2129
```

RESULT 10
CR598188

```
LOCUS       CR598188            1607 bp    mRNA    linear   HTC 21-JUL-2004
DEFINITION  full-length cDNA clone CS0DJ012YI18 of T cells (Jurkat cell line)
            Cot 10-normalized of Homo sapiens (human).
ACCESSION   CR598188
VERSION     CR598188.1  GI:50478995
KEYWORDS    HTC; CNSLT_cDNA.
SOURCE      Homo sapiens (human)
  ORGANISM  Homo sapiens
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Euarchontoglires; Primates; Catarrhini;
            Hominidae; Homo.
REFERENCE   1  (bases 1 to 1607)
  AUTHORS   Li,W.B., Gruber,C., Jessee,J. and Polayes,D.
  TITLE     Full-length cDNA libraries and normalization
  JOURNAL   Unpublished
  REMARK    Contact : Feng Liang Email : fliang@lifetech.com URL :
            http://fulllength.invitrogen.com/ InVitroGen Corporation 1600
            Faraday Avenue
REFERENCE   2  (bases 1 to 1607)
  AUTHORS   Genoscope.
  TITLE     Direct Submission
  JOURNAL   Submitted (20-JUL-2004) Genoscope - Centre National de Sequencage :
            BP 191 91006 EVRY cedex - FRANCE (E-mail : seqref@genoscope.cns.fr
            - Web : www.genoscope.cns.fr)
COMMENT     1st strand cDNA was primed with a NotI-oligo(dT) primer. Five prime
            end enriched, double-strand cDNA was digested with Not I and cloned
            into the Not I and EcoR V sites of the pCMVSPORT 6 vector. Library
            was normalized. Library was constructed by Life Technologies, a
            division of Invitrogen.
FEATURES             Location/Qualifiers
     source          1..1607
                     /organism="Homo sapiens"
                     /mol_type="mRNA"
                     /db_xref="taxon:9606"
                     /clone="CS0DJ012YI18"
                     /tissue_type="T cells (Jurkat cell line) Cot
                     10-normalized"
                     /plasmid="pCMVSPORT_6"
ORIGIN
```

```
Alignment Scores:
Pred. No.:                2.39e-104      Length:         1607
Score:                    1334.00        Matches:        272
Percent Similarity:       70.74%         Conservative:   93
Best Local Similarity:    52.71%         Mismatches:     127
Query Match:              28.94%         Indels:         24
DB:                       4              Gaps:           4
```

US-10-768-511-8 (1-901) x CR598188 (1-1607)

```
Qy        386 HisThrValSerMetCysLeuTyrAsnThrMetValAspAlaValLysProGlyAspArg 405
              ||||||||    :::         :::|||   :::||||||   |||:::|||||||||||
Db          3 CACACAGTTATCCTGTTTGCTCACAATGATCTCGTTGACAAGGTCCAGCCTGGGGACAGA 62

Qy        406 IleGluValThrGlyValPheLysAlaMetAlaValArgValGlyProAsnGlnArgThr 425
              :::   |||||||||:::::::::|||:::    :::||||||    |||
Db         63 GTGAATGTTACAGGCATCTATCGAGCTGTGCCTATTCGAGTCAATCCAAGAGTGAGTAAT 122

Qy        426 LeuArgAlaLeuTyrLysThrTyrIleAspCysValHisValLysLysSerAspArgGly 445
              ::::::::::::|||||||||:::||||||    :::|||     :::|||:::|||
Db        123 GTGAAGTCTGTCTACAAAACCCACATTGATGTCATTCATTATCGGAAAACGGATGCAAAA 182

Qy        446 ArgLeuGlnThrGluAspProMetGluMetAspLysGluAsnAspMetTyrAlaGlyTyr 465
              ||||||                                                ||
Db        183 CGTCTGCAT---------------------------------------------GGCCTT 197

Qy        466 HisGluSerAspThrSerGluAlaAlaAsnGluAlaLysIleGlnLysLeuLysGluLeu 485
                 |||            :::      :::|||   :::::::::   ||||||||||||
Db        198 GATGAAGAAGCAGAACAGAAACTTTTTCAGAGAAACGTGTGGAATTGCTTAAGGAACTT 257

Qy        486 SerLysLeuProGlyIleTyrAspArgLeuSerArgSerLeuAlaProSerIleTrpGlu 505
              |||:::   |||   ||||||:::||||||:::   :::||||||||||||||:::|||
Db        258 TCCAGGAAACCAGACATTTATGAGAGGCTTGCTTCAGCCTTGGCTCCAAGCATTTATGAA 317

Qy        506 LeuGluAspIleLysLysGlyLeuLeuCysGlnLeuPheGlyGlyLysAlaLysLysIle 525
                 |||||||||||||||||:::|||   |||||||||||||||          |||
Db        318 CATGAAGATATAAAGAAGGGAATTTTGCTTCAGCTCTTTGGCGGGACAAGGAAGGATTTT 377

Qy        526 ------ProSerGlyAlaSerPheArgGlyAspIleAsnValLeuLeuValGlyAspPro 543
                    ||||||   :::||||||:::||||||   ||||||||
Db        378 AGTCACACTGGAAGGGGCAAATTTCGGGCTGAGATCAACATCTTGCTGTGTGGCGACCCT 437

Qy        544 GlyThrSerLysSerGlnLeuLeuGlnTyrValHisLysIleAlaProArgGlyIleTyr 563
              ||||||||||||||||||||||||||||||:::   :::   ||||||||    |||
Db        438 GGTACCAGCAAGTCCCAGCTGCTGCAGTACGTGTACAACCTCGTCCCCAGGGGCCAGTAC 497

Qy        564 ThrSerGlyArgGlySerSerAlaValGlyLeuThrAlaTyrValThrLysAspProGlu 583
              |||||||:::|||||||||||||||||||||||||||||||||   ||||||||||||
Db        498 ACGTCTGGGAAGGGCTCCAGTGCAGTTGGCCTCACTGCGTACGTAATGAAAGACCCTGAG 557

Qy        584 ThrArgGluThrValLeuGluSerGlyAlaLeuValLeuSerAspArgGlyIleCysCys 603
              ||||||:::   ||||||::::::||||||||||||||||||||   ||||||||||||
Db        558 ACAAGGCAGCTGGTCCTGCAGACAGGTGCTCTTGTCCTGAGTGACAACGGCATCTGCTGT 617

Qy        604 IleAspGluPheAspLysMetSerAspAsnAlaArgSerMetLeuHisGluValMetGlu 623
              ||||||||||||||||||||||::::::::   ||||||:::|||||||||||||||||
Db        618 ATCGATGAGTTCGACAAGATGAATGAAAGTACAAGATCGGTATTGCATGAAGTCATGGAA 677

Qy        624 GlnGlnThrValSerValAlaLysGlyGlyIleIleAlaSerLeuAsnAlaArgThrSer 643
              ||||||||:::|||:::||||||   |||||||||     ||||||||||||||||||
Db        678 CAGCAGACTCTGTCCATTGCAAAGGCTGGGATCATCTGTCAGCTCAATGCGCGCACCTCT 737

Qy        644 ValLeuAlaCysAlaAsnProSerGlySerArgTyrAsnAlaArgLeuSerValIleAsp 663
              |||||||||   |||||||||      |||:::::::|||   :::   :::   |||:::
```

```
Db        738 GTCCTGGCAGCAGCAAATCCCATTGAGTCTCAGTGGAATCCTAAAAAAACAACCATTGAA 797

Qy        664 AsnIleGlnLeuProProThrLeuLeuSerArgPheAspLeuIleTyrLeuMetLeuAsp 683
Db        798 AACATCCAGCTGCCTCATACTTTATTATCAAGGTTTGATTTGATCTTCCTCTTGCTGGAC 857

Qy        684 LysProAspGluGlnAsnAspArgArgLeuAlaArgHisLeuValAlaLeuHisTyrGlu 703
Db        858 CCTCAGGACGAAGCCTATGACAGGCGTCTGGCTCACCACCTGGTCGCACTGTACTACCAG 917

Qy        704 AsnTyrGluValSerLysGlnAspAlaLeuAspLeuGlnThrLeuThrAlaTyrIleThr 723
Db        918 AGCGAGGAGCAGGCAGAGGAGGAGCTCCTGGACATGGCGGTGCTAAAGGACTACATTGCC 977

Qy        724 TyrAlaArgGlnHisValHisProThrLeuSerAspGluAlaAlaGluAspLeuIleAsn 743
Db        978 TACGCGCACAGCACCATCATGCCGCGGCTAAGTGAGGAAGCCAGCCAGGCTCTCATCGAG 1037

Qy        744 GlyTyrValGluMetArgGlnLysGlyAsnPheProGlySerSerLysLysValIleThr 763
Db       1038 GCTTATGTAGACATGAGGAAG-----------ATTGGCAGTAGCCGGGGAATGGTTTCT 1085

Qy        764 AlaThrProArgGlnLeuGluSerMetIleArgIleSerGluAlaLeuAlaArgMetArg 783
Db       1086 GCATACCCTCGACAGCTAGAGTCATTAATCCGCTTAGCAGAAGCCCATGCTAAAGTAAGA 1145

Qy        784 PheSerGluValValGluLysValAspAlaAlaGluAlaValArgLeuLeuAspValAla 803
Db       1146 TTGTCTAACAAAGTTGAAGCCATTGATGTGGAAGAGGCCAAACGCCTCCATCGGGAAGCT 1205

Qy        804 LeuGlnGlnSerAlaThrAspHisAlaThrGlyThrIleAspMetAspLeuIleThrThr 823
Db       1206 CTGAAGCAGTCTGCAACTGATCCCCGGACTGGCATCGTGGACATATCTATTCTTACTACG 1265

Qy        824 GlyValSerAlaSerGluArgIleArgArgAlaAsnLeuLeuAlaAlaLeuArgGluLeu 843
Db       1266 GGGATGAGTGCCACCTCTCGTAAACGGAAAGAAGAATTAGCTGAAGCATTGAAAAAGCTT 1325

Qy        844 IleAlaAspLysIleSerProGlySerSerSerGlyLeuLysThrSerGlnLeuLeuGlu 863
Db       1326 ATT---------TTATCTAAGGGCAAAACACCAGCTCTAAAATACCAGCAACTTTTTGAA 1376

Qy        864 AspIleArgSerGlnSerSerValAspValSerLeuGlnAspIleLysAsnAlaLeuGly 883
Db       1377 GATATTCGGGGACAATCTGACATAGCAATTACTAAAGATATGTTGAAGAAGCACTGCGT 1436

Qy        884 SerLeuGlnGlyGluGlyPheLeuThrValHisGlyAspIleValLys 899
Db       1437 GCCCTGGCAGATGATGATTTCCTGACAGTGACTGGGAAGACCGTGCGC 1484

RESULT 11
DQ057540
LOCUS       DQ057540                3389 bp    mRNA    linear   HTC 08-AUG-2005
DEFINITION  Antonospora locustae clone 2f18 mRNA sequence.
ACCESSION   DQ057540
VERSION     DQ057540.1  GI:68301743
KEYWORDS    HTC.
SOURCE      Antonospora locustae
  ORGANISM  Antonospora locustae
            Eukaryota; Fungi; Microsporidia; Antonospora.
REFERENCE   1  (bases 1 to 3389)
  AUTHORS   Williams,B.A., Slamovits,C.H., Patron,N.J., Fast,N.M. and
            Keeling,P.J.
  TITLE     A high frequency of overlapping gene expression in compacted
            eukaryotic genomes
  JOURNAL   Proc. Natl. Acad. Sci. U.S.A. 102 (31), 10936-10941 (2005)
   PUBMED   16037215
REFERENCE   2  (bases 1 to 3389)
  AUTHORS   Williams,B.A.P., Slamovits,C.H., Patron,N.J., Fast,N.M. and
            Keeling,P.J.
  TITLE     Direct Submission
  JOURNAL   Submitted (09-MAY-2005) Botany, University of British Columbia,
            3529-6270 University Boulevard, Vancouver, BC V6T 1Z4, Canada
FEATURES             Location/Qualifiers
     source          1..3389
                     /organism="Antonospora locustae"
                     /mol_type="mRNA"
                     /db_xref="taxon:278021"
                     /clone="2f18"
ORIGIN

Alignment Scores:
Pred. No.:                6.18e-101     Length:         3389
Score:                    1300.00       Matches:        299
Percent Similarity:       59.92%        Conservative:   148
Best Local Similarity:    40.08%        Mismatches:     247
Query Match:              28.21%        Indels:         52
DB:                       4             Gaps:           17

US-10-768-511-8 (1-901) x DQ057540 (1-3389)

Qy        142 GlnSerArgGluProGlyHisArgSerProSerArgGluPro--SerAlaAspGlyArgP 161
Db       1025 CAATCAATTAACCCATGGATCCGTCGCAGATCTCGTCAGCCTTTTCCTTTGATGAAGCAA 1084

Qy        161 roSerGluSerAlaGluProAspAspThrLeuGlyGlyGluTyrAla------------- 176
Db       1085 GCCAGGACATAGCGCAACCGCTGAATACA---CCTACAGAGTACAGCTCCGGATACGTCT 1141

Qy        177 -------------------------------TyrValTrpGlyThrAsnValAsnIleP 186
Db       1142 CGCAGCCTCTAATAGAGACAGAAAGAATCAAAGTGATCTGGGGCACGACCATCAACGTGC 1201

Qy        186 roAspValLeuArgAlaIleArgArgPheLeuHisAsnTyrArgSerSerAlaHisAspL 206
Db       1202 AGGAGTCATCGGAAATGTTCAAGGAATTTGTGAGGAGCTTT------GTGGCACACGACA 1255

Qy        206 euAsnSerLysTyrIleGlnIleIleGluGluThrValGluArgGluGluAspThrLeuA 226
Db       1256 CACAC---GTTTACATGGAGAAGATTCGCGCTCTGATTGATACACAGTCAAGTACTCTTC 1312

Qy        226 snIleAspMetSerAspIleTyrAspHisAspProAsp---LeuTyrAlaLysIleValA 245
Db       1313 TGATGAATTGCAAG------------CACCTGCCCGAGATGCTGTATAAGCAACTGGAAT 1360

Qy        245 rgTyrProLeuAspIleIleProLeuLeuAspThrGluCysGlnGluValAlaThrSerL 265
Db       1361 CGTATCCGCAAGAGATTATACCCATTTTTGAGAACTCTCTGAGCGAGGTGGTTTCAGAGC 1420

Qy        265 euLeuProThrPheGluLysHisIleGluAlaArgProPheAsnLeuLysAlaSerValH 285
Db       1421 TTGCGCCAGAGTTCAACTGTGTAGTGCGAATCAGGTTTTTAATATCGGCCGCCAACTGT 1480

Qy        285 isMetArgGluLeuAsnProSerAspIleAspLysLeuValSerValLysGlyMetValI 305
Db       1481 TCATCAGAAACATCGACCCGCGCGACATCGACAGAATTGTTGCGGTGCGAGGCATGGTGA 1540

Qy        305 leArgCysSerSerIleIleProGluIleLysGlyAlaPhePheLysCysLeuValCysG 325
Db       1541 TAAGAGTGTCAAACGTGATGCCTGAAATACTTAAGGCATTTTATACATGCACTCGATGTG 1600

Qy        325 lyHisSerProProLeuValThrValValLysGlyArgValGluGluProThrArgCysG 345
Db       1601 GA---AAGGAAGTACTTGTCGAAAACATACGTGGCGTGATTAACGAGCCCACATCCTGC- 1656

Qy        345 luLysProGluCysAlaAlaArgAsnAlaMetSerLeuIleHisAsnArgCysThrPheA 365
Db       1657 --------GACTGCGGTGCAAGGTTTGCATACCAACTTGTCTCCAACAAAAGCATCTACA 1708

Qy        365 laAsnLysGlnIleValArgLeuGlnGluThrProAspAlaIleProGluGlyGluThrP 385
Db       1709 TAGATAAACAGATCATTCGAGTCCAGGAGCTTCCTGAGAATATACAGGACGGAACTACGC 1768

Qy        385 roHisThrValSerMetCysLeuTyrAsnThrMetValAspAlaValLysProGlyAspA 405
Db       1769 CAATGCCAATAACAGTGCTTGCGCAAGACGATTTAGTAGACACTGTGATACCTGGCGACA 1828
```

```
Qy        405 rgIleGluValThrGlyValPheLysAlaMetAlaValArgValGlyProAsnGlnArgT 425
              ::::::::::|||||:::   :::|||        ||||||:::          :::
Db       1829 AGGTGAAGCTCACAGGCATCATGAGGGCGTCTCCGGTCAGGCTGAATCACTTTTCAAAAA 1888

Qy        425 hrLeuArgAlaLeuTyrLysThrTyrIleAspCysValHisValLysLysSerAspArgG 445
                :::::::::   ::::::   :::|||:::   :::       :::       |||
Db       1889 AGATAAAAAGCAGCTTTCGCGTCTTCATTGAGCTTCTTGCACAGGAAACGATTGACGCAA 1948

Qy        445 lyArgLeuGlnThrGlu------------AspProMetGluMetAspLysGluAsnAspM 461
                       ::::::             |||      |||:::       ::::::
Db       1949 CCGATTTCAAGAGCACTAGATGCCATGGTGACTGCGCAGAAGTCTCTGGTGCAAGCGAAA 2008

Qy        461 etTyrAlaGlyTyrHisGluSerAspThrSerGluAlaAlaAsnGluAlaLysIleGlnL 481
                    |||         |||   :::   |||      :::   :::     |||::::
Db       2009 GCGTTGCTAGAAGCGGTGAA---GAAGCCAGCAGCAAAAGCAGAGACGTGCTGATAGAGG 2065

Qy        481 ysLeuLysGluLeuSerLysLeuProGlyIleTyrAspArgLeuSerArgSerLeuAlaP 501
              ::::::::::|||          |||      ||||||   |||:::      :::||||
Db       2066 AGATTGAGAAGCTGAGAAACAACCCCAGAAGATACGATATTCTTGCTGGTATGATCGCAC 2125

Qy        501 roSerIleTrpGluLeuGluAspIleLysLysGlyLeuLeuCysGlnLeuPheGlyGlyL 521
              |||||:::      :::|||::::::||||||   ||||||   |||||||||||||||
Db       2126 CGTCTGTGTGTGGAATGGAAAACGTGAAGAAGGCGCTTCTGCTACAGTTGTTTGGTGGA- 2184

Qy        521 ysAlaLysLysIleProSerGlyAlaSerPheArgGlyAspIleAsnValLeuLeuValG 541
                 :::|||                     ||||||||||||||:::||||||     |
Db       2185 --GTTCGAAAGGAGGGCGAGAACATGAAGCTTAGGGGCGACATCAACATTCTTCTTGCAG 2242

Qy        541 lyAspProGlyThrSerLysSerGlnLeuLeuGlnTyrValHisLysIleAlaProArgG 561
              ||||||||||   ||||||||||||||||||   ::::::::::::|||      ||||
Db       2243 GAGACCCAGGGATTTCCAAGTCGCAGCTTCTTTCCTTCATCAACAGAATCTGCAGCAGAG 2302

Qy        561 lyIleTyrThrSerGlyArgGlySerSerAlaValGlyLeuThrAlaTyrValThrLysA 581
              ||||||||||||||||||:::||||||||||||||||||:::|||   |||      :::|
Db       2303 GAATCTACACGAGCGGAAAAGGAAGCTCTGCAGTGGGCCTGAGTGCGACTGTCAAACGAG 2362

Qy        581 spProGluThrArgGluThrValLeuGluSerGlyAlaLeuValLeuSerAspArgGlyI 601
              ||   :::|||:::         ||||||||   |||||||||||||||||||   ||||
Db       2363 ATGTGGATACAAAAACGTTTGTTCTTGAACCAGGCGCCCTGGTTCTCAGCGATAATGGCA 2422

Qy        601 leCysCysIleAspGluPheAspLysMetSerAspAsnAlaArgSerMetLeuHisGluV 621
              ||||||||||||||||||||||||||||||||||:::      ||||||:::|||||||||
Db       2423 TCTGCTGCATCGACGAGTTTGACAAGATGAGCGACAGCACGAGAAGCGTACTTCACGAGG 2482

Qy        621 alMetGluGlnGlnThrValSerValAlaLysGlyGlyIleIleAlaSerLeuAsnAlaA 641
              |||||||||||||||||||:::||||||||||   |||||||||   :::|||||||||||
Db       2483 TCATGGAGCAGCAGACCGTGACGGTTGCTAAGGCGGGTATAATAACCACGTTGAATGCAA 2542

Qy        641 rgThrSerValLeuAlaCysAlaAsnProSerGlySerArgTyrAsnAlaArgLeuSerV 661
              ||   ||||||||||||         ||||||   |||:::||||||   :::   |||:
Db       2543 GATGCTCCGTGCTGGCAAGCTGCAATCCAATTGAGAGCAAGTACAACACCAAGAAGAGCA 2602

Qy        661 alIleAspAsnIleGlnLeuProProThrLeuLeuSerArgPheAspLeuIleTyrLeuM 681
              ::||||||||||:::   :::   ||||||||||||||||||||||||||||::::::   |||:
Db       2603 TCATAGACAATCTGCACATAGTCCCCACTCTTTTGAGCCGCTTTGATGTAGTTTGCCTTC 2662

Qy        681 etLeuAspLysProAspGluGlnAsnAspArgArgLeuAlaArgHisLeuValAlaLeuH 701
              :::::||||||   ||||||   :::||||||   :::|||   |||:::::::::
Db       2663 TGATTGACAAGAGCGACGAGGTGAGCGACAGAACAGTTGCAGAACACATACTTTCCTTC- 2721

Qy        701 isTyrGluAsnTyrGluValSerLysGlnAspAlaLeuAspLeuGlnThrLeuThrAlaT 721
                |||::::::   |||      |||      :::      ::::::   |||      |
Db       2722 --TATGATAGTAGAGAGGATGAAAAAACCAGCTCA---AGCATAGAACTGCTGAAAGAAT 2776

Qy        721 yrIleThrTyrAlaArgGlnHisValHisProThrLeuSerAspGluAlaAlaGluAspL 741
              |||||      ||||||:::   ::::::|||   |||:::   |||:::         |
Db       2777 ATATAAAGGAGGCGCGGAAG---ATAAATCCGAAGCTCACAAAAGAAAGTATGTCCGCTC 2833

Qy        741 euIleAsnGlyTyrValGluMetArgGlnLysGlyAsnPheProGlySerSerLysLysV 761
              ||         |||   |||:::||||||         |||                   |||
Db       2834 TGAGAGAGTGTTACTGCGAGCTCCGCCAGCTGGACAAC---------------GGAAAAA 2878

Qy        761 alIleThrAlaThrProArgGlnLeuGluSerMetIleArgIleSerGluAlaLeuAlaA 781
                ||||||||||||   ||||||||||||||:::||||||         ||||||   ||||
Db       2879 GCATCACTGCGACAACGCGCCAGTTAGAAAGCCTGATCAGATTCATCGAGGCACATGCGC 2938

Qy        781 rgMetArgPheSerGluValValGluLysValAspAlaAlaGluAlaValArgLeuLeuA 801
                |||||||   |||      ||||||         |||      ||||||:::|||::::::
Db       2939 GAATGCGGCTGAGCAGCACTGTGGAGGCCGAGGACGTGAATGAGGCGATTCGAATTGTGA 2998

Qy        801 spValAlaLeuGlnGlnSerAlaThrAspHisAlaThrGlyThrIleAspMetAspLeuI 821
                 :::|||         |||   |||         ||||||   ||||||||||||||:
Db       2999 AAGAAAGCCTGTTGATGTACGCAATAGACCCGACAACAGGAAAGATAGACATGGACCTTG 3058

Qy        821 leThrThrGlyValSerAlaSerGluArgIleArgArgAlaAsnLeuLeuAlaAlaLeuA 841
              ::   :::|||   |||                 :::   :::      |||:::      |||
Db       3059 TTATAAGCGGAAAGTCC------------GTGTACAAGAAAAAGCTCATGGAGGACCTCA 3106

Qy        841 rgGluLeuIleAlaAspLysIleSerProGlyProGlySerSerSerGlyLeuLysThrSerGlnL 861
                       :::|||      :::         :::                  |||      |
Db       3107 ACCACGAAGTTGCAAAACTTGTGCGACAGCCAACA------------AAGTTTGTGGATC 3154

Qy        861 euLeuGluAspIle 865
                ||||||||   |||
Db       3155 TTCTTGAAAAGATA 3168
```

```
RESULT 12
CR606514
LOCUS       CR606514                1833 bp    mRNA    linear   HTC 21-JUL-2004
DEFINITION  full-length cDNA clone CS0DK011YN01 of HeLa cells Cot 25-normalized
            of Homo sapiens (human).
ACCESSION   CR606514
VERSION     CR606514.1  GI:50487321
KEYWORDS    HTC; CNSLT_cDNA.
SOURCE      Homo sapiens (human)
  ORGANISM  Homo sapiens
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Euarchontoglires; Primates; Catarrhini;
            Hominidae; Homo.
REFERENCE   1  (bases 1 to 1833)
  AUTHORS   Li,W.B., Gruber,C., Jessee,J. and Polayes,D.
  TITLE     Full-length cDNA libraries and normalization
  JOURNAL   Unpublished
  REMARK    Contact : Feng Liang Email : fliang@lifetech.com URL :
            http://fulllength.invitrogen.com/ InVitroGen Corporation 1600
            Faraday Avenue
REFERENCE   2  (bases 1 to 1833)
  AUTHORS   Genoscope.
  TITLE     Direct Submission
  JOURNAL   Submitted (20-JUL-2004) Genoscope - Centre National de Sequencage :
            BP 191 91006 EVRY cedex - FRANCE (E-mail : seqref@genoscope.cns.fr
            - Web : www.genoscope.cns.fr)
COMMENT     1st strand cDNA was primed with a NotI-oligo(dT) primer. Five prime
            end enriched, double-strand cDNA was digested with Not I and cloned
            into the Not I and EcoR V sites of the pCMVSPORT 6 vector. Library
            was normalized. Library was constructed by Life Technologies, a
            division of Invitrogen.
FEATURES             Location/Qualifiers
     source          1..1833
                     /organism="Homo sapiens"
                     /mol_type="mRNA"
                     /db_xref="taxon:9606"
                     /clone="CS0DK011YN01"
                     /tissue_type="HeLa cells Cot 25-normalized"
                     /plasmid="pCMVSPORT_6"
ORIGIN

Alignment Scores:
Pred. No.:               3.59e-97      Length:         1833
Score:                   1252.00       Matches:        258
Percent Similarity:      70.41%        Conservative:   87
Best Local Similarity:   52.65%        Mismatches:     121
Query Match:             27.16%        Indels:         24
DB:                      4             Gaps:           4
```

```
US-10-768-511-8 (1-901) x CR606514 (1-1833)

Qy        412 PheLysAlaMetAlaValArgValGlyProAsnGlnArgThrLeuArgAlaLeuTyrLys 431
              ::::::|||:::    :::|||||   |||             :::::::::::::||||||
Db          1 TATCGAGCTGTGCCTATTCGAGTCAATCCAAGAGTGAGTAATGTGAAGTCTGTCTACAAA 60

Qy        432 ThrTyrIleAspCysValHisValLysLysSerAspArgGlyArgLeuGlnThrGluAsp 451
              |||:::||||||    :::|||    :::|||:::|||         ||||||
Db         61 ACCCACATTGATGTCATTCATTATCGGAAAACGGATGCAAAACGTCTGCAT--------- 111

Qy        452 ProMetGluMetAspLysGluAsnAspMetTyrAlaGlyTyrHisGluSerAspThrSer 471
                                                   |||      |||
Db        112 ------------------------------------GGCCTTGATGAAGAAGCAGAACAG 135

Qy        472 GluAlaAlaAsnGluAlaLysIleGlnLysLeuLysGluLeuSerLysLeuProGlyIle 491
              :::       :::|||    :::::::::    ||||||||||||||:::     |||   |||
Db        136 AAACTTTTTTCAGAGAAACGTGTGGAATTGCTTAAGGAACTTTCCAGGAAACCAGACATT 195

Qy        492 TyrAspArgLeuSerArgSerLeuAlaProSerIleTrpGluLeuGluAspIleLysLys 511
              |||:::||||||:::    :::|||||||||||||||:::|||   ||||||||||||||
Db        196 TATGAGAGGCTTGCTTCAGCCTTGGCTCCAAGCATTTATGAACATGAAGATATAAAGAAG 255

Qy        512 GlyLeuLeuCysGlnLeuPheGlyGlyLysAlaLysLysIle------ProSerGlyAla 529
              |||:::|||    |||||||||||||||            |||
Db        256 GGAATTTTGCTTCAGCTCTTTGGCGGGACAAGGAAGGATTTTAGTCACACTGGAAGGGGC 315

Qy        530 SerPheArgGlyAspIleAsnValLeuLeuValGlyAspProGlyThrSerLysSerGln 549
                 ||||||    :::||||||:::||||||   ||||||||||||||||||||||||||||
Db        316 AAATTTCGGGCTGAGATCAACATCTTGCTGTGTGGCGACCCTGGTACCAGCAAGTCCCAG 375

Qy        550 LeuLeuGlnTyrValHisLysIleAlaProArgGlyIleTyrThrSerGlyArgGlySer 569
              ||||||||||||||||:::    :::      ||||||||   ||||||||||||:::||||||
Db        376 CTGCTGCAGTACGTGTACAACCTCGTCCCCAGGGGCCAGTACACGTCTGGGAAGGGCTCC 435

Qy        570 SerAlaValGlyLeuThrAlaTyrValThrLysAspProGluThrArgGluThrValLeu 589
              ||||||||||||||||||||||||||||   ||||||||||||||||:::   ||||||
Db        436 AGTGCAGTTGGCCTCACTGCGTACGTAATGAAAGACCCTGAGACAAGGCAGCTGGTCCTG 495

Qy        590 GluSerGlyAlaLeuValLeuSerAspArgGlyIleCysCysIleAspGluPheAspLys 609
              ::::::||||||||||||||||||||||||   ||||||||||||||||||||||||||||
Db        496 CAGACAGGTGCTCTTGTCCTGAGTGACAACGGCATCTGCTGTATCGATGAGTTCGACAAG 555

Qy        610 MetSerAspAsnAlaArgSerMetLeuHisGluValMetGluGlnGlnThrValSerVal 629
              |||:::::::::      ||||||:::||||||||||||||||||||||||:::|||:::
Db        556 ATGAATGAAAGTACAAGATCGGTATTGCATGAAGTCATGGAACAGCAGACTCTGTCCATT 615

Qy        630 AlaLysGlyGlyIleIleAlaSerLeuAsnAlaArgThrSerValLeuAlaCysAlaAsn 649
              ||||||   |||||||||      |||||||||||||||||||||||||||   ||||||
Db        616 GCAAAGGCTGGGATCATCTGTCAGCTCAATGCGCGCACCTCTGTCCTGGCAGCAGCAAAT 675

Qy        650 ProSerGlySerArgTyrAsnAlaArgLeuSerValIleAspAsnIleGlnLeuProPro 669
              |||       |||::::::|||      :::      :::      |||:::|||||||||||||||
Db        676 CCCATTGAGTCTCAGTGGAATCCTAAAAAAACAACCATTGAAAACATCCAGCTGCCTCAT 735

Qy        670 ThrLeuLeuSerArgPheAspLeuIleTyrLeuMetLeuAspLysProAspGluGlnAsn 689
              ||||||||||||||||||||||||||||:::|||:::||||||         ||||||
Db        736 ACTTTATTATCAAGGTTTGATTTGATCTTCCTCTTGCTGGACCCTCAGGACGAAGCCTAT 795

Qy        690 AspArgArgLeuAlaArgHisLeuValAlaLeuHisTyrGluAsnTyrGluValSerLys 709
              ||||||||||||||   |||||||||||||||:::|||::::::    |||     ::::::
Db        796 GACAGGCGTCTGGCTCACCACCTGGTCGCACTGTACTACCAGAGCGAGGAGCAGGCAGAG 855

Qy        710 GlnAspAlaLeuAspLeuGlnThrLeuThrAlaTyrIleThrTyrAlaArgGlnHisVal 729
              ::::::    ||||||:::      |||         ||||||   ||||||         :::
Db        856 GAGGAGCTCCTGGACATGGCGGTGCTAAAGGACTACATTGCCTACGCGCACAGCACCATC 915

Qy        730 HisProThrLeuSerAspGluAlaAlaGluAspLeuIleAsnGlyTyrValGluMetArg 749
                 |||      ||||||:::|||||||||::::::      ||||||      ||||||:::|||
Db        916 ATGCCGCGGCTAAGTGAGGAAGCCAGCCAGGCTCTCATCGAGGCTTATGTAGACATGAGG 975

Qy        750 GlnLysGlyAsnPheProGlySerSerLysLysValIleThrAlaThrProArgGlnLeu 769
              :::            |||||||||:::   :::::::::|||   ||||||||||||
Db        976 AAG-----------ATTGGCAGTAGCCGGGGAATGGTTTCTGCATACCCTCGACAGCTA 1023

Qy        770 GluSerMetIleArgIleSerGluAlaLeuAlaArgMetArgPheSerGluValValGlu 789
              ||||||:::||||||::::::||||||      |||::::::|||   |||         ||||||
Db       1024 GAGTCATTAATCCGCTTAGCAGAAGCCCATGCTAAAGTAAGATTGTCTAACAAAGTTGAA 1083

Qy        790 LysValAspAlaAlaGluAlaValArgLeuLeuAspValAlaLeuGlnGlnSerAlaThr 809
                 :::|||      ||||||   ||||||            ||||||:::||||||||||||
Db       1084 GCCATTGATGTGGAAGAGGCCAAACGCCTCCATCGGGAAGCTCTGAAGCAGTCTGCAACT 1143

Qy        810 AspHisAlaThrGlyThrIleAspMetAspLeuIleThrThrGlyValSerAlaSerGlu 829
              |||      ||||||      :::|||:::      ::::::|||||||||:::||||||:::
Db       1144 GATCCCCGGACTGGCATCGTGGACATATCTATTCTTACTACGGGGATGAGTGCCACCTCT 1203

Qy        830 ArgIleArgArgAlaAsnLeuLeuAlaAlaLeuArgGluLeuIleAlaAspLysIleSer 849
              |||   |||:::      |||      ||||||::::::||||||         :::|||
Db       1204 CGTAAACGGAAAGAAGAATTAGCTGAAGCATTGAAAAAGCTTATT---------TTATCT 1254

Qy        850 ProGlySerSerSerGlyLeuLysThrSerGlnLeuLeuGluAspIleArgSerGlnSer 869
                 |||   :::      ||||||      ||||||   |||||||||||   ||||||
Db       1255 AAGGGCAAAACACCAGCTCTAAAATACCAGCAACTTTTTGAAGATATTCGGGGACAATCT 1314

Qy        870 SerValAspValSerLeuGlnAspIleLysAsnAlaLeuGlySerLeuGlnGlyGluGly 889
                 :::   ::::::            :::   ||||||   :::|||      :::
Db       1315 GACATAGCAATTACTAAAGATATGTTTGAAGAAGCACTGCGTGCCCTGGCAGATGATGAT 1374

Qy        890 PheLeuThrValHisGlyAspIleValLys 899
              ||||||||||||   |||      |||:::
Db       1375 TTCCTGACAGTGACTGGGAAGACCGTGCGC 1404
```

```
RESULT 13
DR495517/c
LOCUS       DR495517                 879 bp    mRNA    linear   EST 06-JUL-2005
DEFINITION  WS0298.BR_E02 SS-IB-A-FL-15 Picea sitchensis cDNA clone WS0298_E02
            3, mRNA sequence.
ACCESSION   DR495517
VERSION     DR495517.1  GI:69453834
KEYWORDS    EST.
SOURCE      Picea sitchensis (Sitka spruce)
  ORGANISM  Picea sitchensis
            Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
            Spermatophyta; Coniferopsida; Coniferales; Pinaceae; Picea.
REFERENCE   1  (bases 1 to 879)
  AUTHORS   Ralph,S., Kolosova,N., Oddy,C., Cooper,D., Butterfield,Y.,
            Kirkpatrick,R., Liu,J., Palmquist,D., Stott,J., Barber,S., Yang,G.,
            Babakaiff,R., Brown-John,M., Chand,S., Featherstone,R., Masson,A.,
            Mayo,M., Moran,J., Olson,T., Wong,D., Friedmann,M.F., Ritland,C.E.,
            Siddiqui,A., Holt,R., Jones,S., Marra,M., Ellis,B.E., Douglas,C.,
            Ritland,K. and Bohlmann,J.
  TITLE     The spruce transcriptome: Analysis of expressed sequence tags from
            multiple cDNA libraries
  JOURNAL   Unpublished (2005)
COMMENT     Contact: Joerg Bohlmann
            Genome BC forest genomics program
            University of British Columbia
            Michael Smith Laboratories, 6174 University Boulevard, Rm. 237,
            Vancouver, British Columbia, Canada, V6T 1Z3
            Tel: 1-604-822-0282
            Fax: 1-604-822-2114
            Email: bohlmann@msl.ubc.ca
            Plate: WS0298  row: E  column: 02
            High quality sequence stop: 879.
FEATURES             Location/Qualifiers
     source          1..879
                     /organism="Picea sitchensis"
                     /mol_type="mRNA"
                     /cultivar="FB3-425"
                     /db_xref="taxon:3332"
                     /clone="WS0298_E02"
                     /sex="Hermaphrodite"
                     /lab_host="E. coli DH10B cells"
                     /clone_lib="SS-IB-A-FL-15"
                     /note="Organ: Bark (with phloem and cambium attached) from
                     two year old clonal trees grown under greenhouse
```

```
                  conditions in standard potting soil mixture; Vector:
                  pBluescript II SK (+) XR; Site_1: XhoI (5' end of cDNA);
                  Site_2: BamHI (3' end of cDNA); Adult white pine weevils
                  (Pissodes strobi) were caged with mesh bags on sapling
                  trees. Bark tissue with phloem and cambium attached was
                  harvested 2 hours, 6 hours and 48 hours after continuous
                  feeding by the insects. mRNA was isolated from each tissue
                  source independently and equal quantities of mRNA from
                  each tissue were then pooled. cDNA was prepared from 20
                  micrograms of mRNA according to the full-length cDNA
                  library construction method described by Carninci P. et
                  al. (2000), Genome Research 10(10):1617-1630 and
                  directionally ligated into the pBluescript II SK (+) XR
                  vector digested with XhoI (5' end) and BamHI (3'). Plasmid
                  DNA was then transformed by electroporation into DH10B
                  cells (Invitrogen) for propagation."
ORIGIN

Alignment Scores:
Pred. No.:              2.34e-92        Length:         879
Score:                  1191.00         Matches:        216
Percent Similarity:     89.73%          Conservative:   46
Best Local Similarity:  73.97%          Mismatches:     30
Query Match:            25.84%          Indels:         0
DB:                     8               Gaps:           0

US-10-768-511-8 (1-901) x DR495517 (1-879)

Qy        288 GluLeuAsnProSerAspIleAspLysLeuValSerValLysGlyMetValIleArgCys 307
Db        878 GAACTGAATCCATCAGATATTGAGAAAATGGTTTCAGTGAAAGGGATGATCATACGATGC 819

Qy        308 SerSerIleIleProGluIleLysGlyAlaPhePheLysCysLeuValCysGlyHisSer 327
Db        818 AGCTCAATAATACCTGAAATCAAAGAGGCATTTTTCAGATGCCTTGTTTGTGGCTATGCT 759

Qy        328 ProProLeuValThrValValLysGlyArgValGluGluProThrArgCysGluLysPro 347
Db        758 CCACCACCAGTGCCTGTTGATAGAGGTCGAATTGATGAACCTACTCGATGTGGAAGGCCA 699

Qy        348 GluCysAlaAlaArgAsnAlaMetSerLeuIleHisAsnArgCysThrPheAlaAsnLys 367
Db        698 GAATGTGCTGCTATGAACTCCATGACTTTAGTTCATAATAGATGCAGGTTTGCTGACAAA 639

Qy        368 GlnIleValArgLeuGlnGluThrProAspAlaIleProGluGlyGluThrProHisThr 387
Db        638 CAGATCGTAAGGCTGCAAGAAACCCCTGATGCTATTCCAGAGGGAGAAACCCCACATACA 579

Qy        388 ValSerMetCysLeuTyrAsnThrMetValAspAlaValLysProGlyAspArgIleGlu 407
Db        578 GTCAGCTTGTTAATGCATGACAAACTAGTTGATGCTGCCAAGCCTGGTGACAGAGTTGAG 519

Qy        408 ValThrGlyValPheLysAlaMetAlaValArgValGlyProAsnGlnArgThrLeuArg 427
Db        518 GTAACAGGAATCTTCAGGGCCATGAGTGTTAGAGTAGGACCAAACCAGAGGACAGTCAAA 459

Qy        428 AlaLeuTyrLysThrTyrIleAspCysValHisValLysLysSerAspArgGlyArgLeu 447
Db        458 TCCCTCTTTAAGACCTACATTGACTGCCTTCACTTGAAGAAGGCAGACAAGACACGACTG 399

Qy        448 GlnThrGluAspProMetGluMetAspLysGluAsnAspMetTyrAlaGlyTyrHisGlu 467
Db        398 CAAGCTGAAGATCCAATGGAGCGTGATGCCACTAGTGACAGCTTTGCTTCTTTTCAAGAG 339

Qy        468 SerAspThrSerGluAlaAlaAsnGluAlaLysIleGlnLysLeuLysGluLeuSerLys 487
Db        338 GGTGATATTTCACCCGTTCAGTATGAAGCCAAGATTGAGAAGTTGAAAGAGCTGTCACGG 279

Qy        488 LeuProGlyIleTyrAspArgLeuSerArgSerLeuAlaProSerIleTrpGluLeuGlu 507
Db        278 CAGCCTGATATATATGAGAAGTTAACCAGGTCATTGGCTCCAAGTATTTGGGAGCTGGAA 219

Qy        508 AspIleLysLysGlyLeuLeuCysGlnLeuPheGlyGlyLysAlaLysLysIleProSer 527
Db        218 GATGTGAAAAAAGGGCTTTTATGTCAGCTTTTGGGGGAACTCCAGTAAAGCTAGCCACA 159

Qy        528 GlyAlaSerPheArgGlyAspIleAsnValLeuLeuValGlyAspProGlyThrSerLys 547
Db        158 GGTGCCACCTTCCGTGGTGACATCAACATTTTGCTTGTTGGGGACCCGGGTACGAGTAAG 99

Qy        548 SerGlnLeuLeuGlnTyrValHisLysIleAlaProArgGlyIleTyrThrSerGlyArg 567
Db         98 TCTCAACTACTTCAGTATGTACACAAGTTGGCTCCTCGTGGTATTTATACAAGTGGTCGG 39

Qy        568 GlySerSerAlaValGlyLeuThrAlaTyrValThr 579
Db         38 GGAAGCTCTGCAGTTGGTTTGACTGCTTATGTGTCA 3

RESULT 14
CO123277
LOCUS       CO123277               881 bp    mRNA    linear   EST 16-JUN-2004
DEFINITION  GR__Eb05F09.f GR__Eb Gossypium raimondii cDNA clone GR__Eb05F09 5',
            mRNA sequence.
ACCESSION   CO123277
VERSION     CO123277.1  GI:48821964
KEYWORDS    EST.
SOURCE      Gossypium raimondii
  ORGANISM  Gossypium raimondii
            Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
            Spermatophyta; Magnoliophyta; eudicotyledons; core eudicotyledons;
            rosids; eurosids II; Malvales; Malvaceae; Malvoideae; Gossypium.
REFERENCE   1  (bases 1 to 881)
  AUTHORS   Kim,H., Yu,Y., Kudrna,D., Hatfield,J., Stum,D., Mueller,C.,
            Udall,J.A., Rapp,R.A., Wendel,J.F., Rao,K., Soderlund,C. and
            Wing,R.A.
  TITLE     Global assembly of Cotton ESTs
  JOURNAL   Unpublished (2004)
COMMENT     Contact: Rod A. Wing
            Arizona Genomics Institute
            The University of Arizona
            Forbes Building Room 303, Tucson, AZ, 85721-0036, USA
            Tel: 520 626 9595
            Fax: 520 621 1259
            Email: rwing@genome.arizona.edu
            Plate: 05  row: F  column: 09.
FEATURES             Location/Qualifiers
     source          1..881
                     /organism="Gossypium raimondii"
                     /mol_type="mRNA"
                     /db_xref="taxon:29730"
                     /clone="GR__Eb05F09"
                     /tissue_type="floral"
                     /dev_stage="_3 to +3 DPA"
                     /lab_host="DH10B"
                     /clone_lib="GR__Eb"
                     /note="Vector: pCMV.SPORT-6.1; Site_1: NotI; Site_2:
                     EcoRV; Library made by Invitrogen with RNA supplied by
                     Wendle lab. Directional cloned into NotI-EV. Colonies
                     plated/picked by AGI. More glycerol clones held in -80."
ORIGIN

Alignment Scores:
Pred. No.:              4.16e-90        Length:         881
Score:                  1165.00         Matches:        230
Percent Similarity:     88.85%          Conservative:   33
Best Local Similarity:  77.70%          Mismatches:     29
Query Match:            25.28%          Indels:         4
DB:                     7               Gaps:           1

US-10-768-511-8 (1-901) x CO123277 (1-881)

Qy        568 GlySerSerAlaValGlyLeuThrAlaTyrValThrLysAspProGluThrArgGluThr 587
Db          1 GGGAGCTCTGCTGTTGGATTGACTGCTTATGTTGCTAAGGACCCTGAAACCGGAGAAACT 60

Qy        588 ValLeuGluSerGlyAlaLeuValLeuSerAspArgGlyIleCysCysIleAspGluPhe 607
```

```
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db         61 GTTTTGGAGAGTGGAGCCCTTGTTCTGAGTGACAGAGGCATCTGTTGTATTGATGAATTT 120

Qy        608 AspLysMetSerAspAsnAlaArgSerMetLeuHisGluValMetGluGlnGlnThrVal 627
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        121 GATAAAATGTCTGATAATGCAAGGAGCATGCTGCATGAGGTCATGGAACAACAGACTGTT 180

Qy        628 SerValAlaLysGlyGlyIleIleAlaSerLeuAsnAlaArgThrSerValLeuAlaCys 647
              |||:::||||||   ||||||||||||||||||||||||||||||||||||||||||||||
Db        181 TCAATAGCAAAAGCTGGTATAATTGCATCTCTTAATGCAAGGACTTCAGTCTTGGCCTGT 240

Qy        648 AlaAsnProSerGlySerArgTyrAsnAlaArgLeuSerValIleAspAsnIleGlnLeu 667
              ||||||||||||||||||||||||||||   ||||||||||||||||||||||   |||
Db        241 GCAAATCCTAGTGGTTCGCGCTATAATCCGCGTCTGTCTGTGATTGACAACATACACCTG 300

Qy        668 ProProThrLeuLeuSerArgPheAspLeuIleTyrLeuMetLeuAspLysProAspGlu 687
              ||||||||||||||||||||||||||||||||||||||:::||||||||   ||||||
Db        301 CCTCCAACCTTGCTGTCCAGATTTGATTTAATCTACTTGATTCTCGACAAGGCTGACGAG 360

Qy        688 GlnAsnAspArgArgLeuAlaArgHisLeuValAlaLeuHisTyrGluAsnTyrGluVal 707
              |||   ||||||||||||||:::|||:::|||:::||||||:::|||:::    |||:::
Db        361 CAGACAGATAGGCGTTTAGCTAAGCATATTGTTTCTTTACATTTTGAGGACACAGAGATT 420

Qy        708 SerLysGlnAspAlaLeuAspLeuGlnThrLeuThrAlaTyrIleThrTyrAlaArgGln 727
              ::::::   |||     :::||||||   |||||||||||||::::::|||||||||:::
Db        421 GCTGAGCATGATGTCATAGATCTTGCTACATTAACAGCATATGTGAGCTATGCTCGCAAG 480

Qy        728 HisValHisProThrLeuSerAspGluAlaAlaGluAspLeuIleAsnGlyTyrValGlu 747
              ::::::||||||   |||||||||||||||||||||:::|||      ||||||||||||
Db        481 AATATTCACCCAAAATTATCTGATGAAGCTGCTGAAGAGCTAACCCGAGGATATGTTGAA 540

Qy        748 MetArgGlnLysGlyAsnPheProGlySerSerLysLysValIleThrAlaThrProArg 767
              ||||||::::::||||||||||||||||||||||||||||||||||||||||||||||||
Db        541 ATGAGGAAAAGAGGAAATTTCCCAGGCAGCAGTAAAAAGGTCATCACAGCAACACCAAGG 600

Qy        768 GlnLeuGluSerMetIleArgIleSerGluAlaLeuAlaArgMetArgPheSerGluVal 787
              |||:::||||||:::||||||:::||||||||||||||||||:::||||||||||||
Db        601 CAAATTGAAAGTCTGATACGCCTTAGTGAAGCCTTGGCTCGTATTCGATTCTCAGAAACA 660

Qy        788 ValGluLysValAspAlaAlaGluAlaValArgLeuLeuAspValAlaLeuGlnGlnSer 807
              ||||||      |||      ||||||   |||||||||:::||||||:::|||||||||
Db        661 GTTGAAATGCGAGATGTAACTGAGGCATTCCGGCTTTTGGAAGTTGCTATGCAGCAATCA 720

Qy        808 AlaThrAspHisAlaThrGlyThrIleAspMetAspLeuIleThrThrGlyValSerAla 827
              ||||||||||||:::|||||||||||||||||||||||||||||||||||||||||||||
Db        721 GCAACTGATCACTCCACAGGAACTATTGATATGGATCTTATCACCACTGGAGTTTCTGCA 780

Qy        828 SerGluArgIleArgArgAlaAsnLeuLeuAlaAlaLeuArgGluLeuIleAlaAspLys 847
              |||||||||:::||||||   ::::::|||:::|||   |||   :::|||   ||||||
Db        781 AGTGAAAGGATGAGGAGGGAGAGTGTGTTATCAGCAACACGCAACGTAATAATGGATAAG 840

Qy        848 IleSerProGlySerSerSerGlyLeuLysThrSerGlnLeuLeuGlu 863
              :::      |||      |||            :::|||||||||
Db        841 CTACAACTTGGGGGACCATCA------------ATGCGATTGCTAGAG 876
```

RESULT 15
CK210237

```
LOCUS       CK210237                1156 bp    mRNA    linear   EST 08-DEC-2003
DEFINITION  FGAS022036 Triticum aestivum FGAS: Library 5 GATE 7 Triticum
            aestivum cDNA, mRNA sequence.
ACCESSION   CK210237
VERSION     CK210237.1  GI:39572627
KEYWORDS    EST.
SOURCE      Triticum aestivum (bread wheat)
  ORGANISM  Triticum aestivum
            Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
            Spermatophyta; Magnoliophyta; Liliopsida; Poales; Poaceae;
            Pooideae; Triticeae; Triticum.
REFERENCE   1  (bases 1 to 1156)
  AUTHORS   Allard,F., Crosby,W.L., Danyluk,J., Eudes,F., Frick,M., Gaudet,D.,
            Genswein,B., Graf,R., Gulick,P., Hrycan,L.D., Laroche,A.,
            Links,M.G., McCarthy,E.L., Monroy,A., Muzak,I., Nilson,D.,
            Penniket,C., Roach,J.L. and Sarhan,F.
  TITLE     Functional Genomics of Abiotic Stress In Wheat and Canola Crops
  JOURNAL   Unpublished (2003)
COMMENT     Contact: Wm L Crosby
            Bioinformatics
            University of Saskatchewan, Department of Computer Science
            1C101 Engineering Building, 57 Campus Drive, Saskatoon,
            Saskatchewan, S7N 5A9, Canada
            Tel: 306 966 1769
            Fax: 306 966 2033
            Email: fgas_ests@cs.usask.ca
            This sequence is the direct result of the Base calling software
            Phred (default parameters). It is the raw base calls. To aid in the
            identification of the high quality insert the software Lucy
            (default parameters) has been run on this sequence. Lucy identified
            the region [16,727].
            Plate: L5B022  row: D  column: 10.
FEATURES             Location/Qualifiers
     source          1..1156
                     /organism="Triticum aestivum"
                     /mol_type="mRNA"
                     /db_xref="taxon:4565"
                     /clone_lib="Triticum aestivum FGAS: Library 5 GATE 7"
                     /note="Vector: pCMV.SPORT6; Crown and developmental stages
                     of spike formation in wheat cultivar Norstar. 4 mRNA
                     populations were combined before constructing the library.
                     The first mRNA population is from 1cm crown sections after
                     30 days of cold acclimation. The second is from 1cm crown
                     sections after 11 days of deacclimation (before
                     deacclimation plants were fully vernalized for 49 days).
                     The third is from different developmental stages of spike
                     formation (5 to 50mm) that still have not emerged from the
                     leaf (dissection required). The last is from different
                     developmental stages of spike and seed formation after
                     having emerged from the leaf (visible). First strand
                     synthesis in this library was done in the presence of
                     methylated dCTP thereby protecting from internal cleavage
                     with NotI."
ORIGIN

Alignment Scores:
Pred. No.:               1.2e-86        Length:          1156
Score:                   1127.00        Matches:         232
Percent Similarity:      84.98%         Conservative:    34
Best Local Similarity:   74.12%         Mismatches:      46
Query Match:             24.45%         Indels:          2
DB:                      7              Gaps:            0

US-10-768-511-8 (1-901) x CK210237 (1-1156)

Qy        536 AsnValLeuLeuValGlyAspProGlyThrSerLysSerGlnLeuLeuGlnTyrValHis 555
              ::::::||||||||||||||||||||||||||||||||||||   ||||||||||:::|||
Db         35 CATATTTTGCTTGTTGGTGATCCTGGAACGAGCAAATCCCATCTTCTCCAGTACATGCAT 94

Qy        556 LysIleAlaProArgGlyIleTyrThrSerGlyArgGlySerSerAlaValGlyLeuThr 575
              |||::::::|||||||||||||||||||||||||||||||||||||||||||||||||||
Db         95 AAACTGTCTCCTCGTGGTATTTACACAAGTGGAAGAGGCAGTTCAGCAGTTGGCCTTACT 154

Qy        576 AlaTyrValThrLysAspProGluThrArgGluThrValLeuGluSerGlyAlaLeuVal 595
              |||||||||   ||||||||||||||||||   ||||||||||||||||||||||||||||
Db        155 GCTTATGTTGCTAAGGACCCTGAAACTGGTGAAACTGTTCTTGAAAGTGGAGCACTTGTT 214

Qy        596 LeuSerAspArgGlyIleCysCysIleAspGluPheAspLysMetSerAspAsnAlaArg 615
              |||||||||:::|||:::|||||||||||||||||||||||||||||||||||||||||||
Db        215 TTGAGTGACAAAGGTGTTTGCTGTATTGATGAGTTTGATAAGATGTCTGATAATGCCCGA 274

Qy        616 SerMetLeuHisGluValMetGluGlnGlnThrValSerValAlaLysGlyGlyIleIle 635
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        275 AGCATGCTGCATGAGGTGATGGAGCAGCAGACTGTATCCATTGCAAAGGCTGGAATTATT 334

Qy        636 AlaSerLeuAsnAlaArgThrSerValLeuAlaCysAlaAsnProSerGlySerArgTyr 655
              ||||||||||||||||||||||||||||||||||||:::||||||   |||||||||
Db        335 GCATCTTTGAATGCTAGGACATCTGTACTAGCATGTGCCAATCCATCTGAATCACGTTAC 394
```

```
Qy       656 AsnAlaArgLeuSerValIleAspAsnIleGlnLeuProProThrLeuLeuSerArgPhe 675
             |||   |||||||||||||||||||||||    ||||||| ||||||||||||||||
Db       395 AATCCAAGGCTTTCTGTGATCGACAATATCCACCTTCCTCCAACACTGCTGTCAAGGTTT 454

Qy       676 AspLeuIleTyrLeuMetLeuAspLysProAspGluGlnAsnAspArgArgLeuAlaArg 695
             |||||||||||||||:::|||||||||   |||||||||   ||||||||||||||:::
Db       455 GACCTGATTTATCTGATCTTGGACAAGGCAGACGAACAAACTGATAGACGCCTGGCTAAG 514

Qy       696 HisLeuValAlaLeuHisTyrGluAsnTyrGluValSerLysGlnAspAlaLeuAspLeu 715
             |||:::|||:::||||||:::||||||   ||||||   :::         |||||||||
Db       515 CATATTGTTTCATTGCATTTCGAGAATCCAGAAGTAGTTGAGCACCAGGTCTTGGATTTG 574

Qy       716 GlnThrLeuThrAlaTyrIleThrTyrAlaArgGlnHisValHisProThrLeuSerAsp 735
                ||||||   |||||||||:::||||||||:::::::::   |||   |||||||||
Db       575 CCCACGTTAGTTGCGTACATCAGCTATGCAAGGAAGTACATTCAGCCAAAGTTATCTGAT 634

Qy       736 GluAlaAlaGluAspLeuIleAsnGlyTyrValGluMetArgGlnLysGlyAsnPhePro 755
             ||||||||||||:::|||      |||||||||   |||||||||:::||||||   |||
Db       635 GAAGCTGCAGAAGAATTGACCCGTGGCTACGTTGCAATGAGGCAAAGGGGGAACAATCCT 694

Qy       756 GlySerSerLysLysValIleThrAlaThrProArgGlnLeuGluSerMetIleArgIle 775
             ||||||   |||||||||||||||||||||   |||   :::||||||:::||||||:::
Db       695 GGTAGCAGAAAGAAGGTCATCACAGCAACAGCTAGGCCAATTGAGAGCTTGATTCGTCTT 754

Qy       776 SerGluAlaLeuAlaArgMetArgPheSerGluValValGluLysValAspAlaAlaGlu 795
             ||||||||||||||||||:::|||   |||   ::::::||||||            :::
Db       755 AGTGAAGCACTGGCGCAAATGCTTTTTCAGAAGTTGGTGGAGTGCGAGATGTAATATCAA 814

Qy       796 AlaValArgLeuLeuAspValAlaLeuGlnGlnSerAlaThrAspHisAlaThrGlyThr 815
             |||:::      |||:::||||||:::|||||||||   |||'||||||||||||||||
Db       815 GCTTTAGGCTTCTTGGAAGTCGCCATGCAGCAATCTTGAACCGATCATGCAACAGGGACG 874

Qy       816 -IleAspMetAspLeuIleThrThrGlyValSerAlaSerGluArgIleArgArgAlaAs 835
               ||||||||||||   |||   ||||||||||||         :::|||   :::|||
Db       875 AATTGATATGGATTCTATCATGACTGGGGTATCCCAAGGTAAAAGGCG-AAACGTGAAAT 933

Qy       835 nLeuLeuAlaAlaLeuArgGluLeuIleAlaAspLys 847
               ||||||   |||:::|||   |||:::      |||
Db       934 CTTGTTGGGGGCTATCCGAAACTTGTTATGGAAAAAA 970
```

Search completed: December 6, 2005, 13:58:12
Job time : 5783 secs